Journal of Quantitative Economics
https://doi.org/10.1007/s40953-019-00184-w

ORIGINAL ARTICLE

Simplified Matrix Methods for Multivariate Edgeworth Expansions

Gubhinder Kundhi[1] · Paul Rilstone[2]

Abstract
Simplified matrix methods are used to analyze the higher order asymptotic properties of $k \times 1$ sample averages. Kronecker differentiation is used to define $k^j \times 1$, j'th order moments, μ_j, cumulants κ_j and Hermite polynomials, H_j. These are then used to derive valid multivariate Edgeworth expansions of arbitrary order having the same form as the standard univariate case: $p(x) = \phi(x)[1+N^{-1/2}\kappa_3' H_3(x)/6+N^{-1}(3\kappa_4' H_4(x)+\kappa_3'^{\otimes 2} H_6(x))/72 + \cdots]$. All the usual steps in the development of a valid Edgeworth expansion are shown to be easily derived using matrix algebra.

Keywords Higher order asymptotics · Edgeworth expansions · Higher order expansions

Introduction

Edgeworth expansions are a central component of higher order asymptotic theory. In themselves, Edgeworth expansions are informative about the sampling distribution of averages and functions of averages including estimators and test statistics. They can be used to improve inferences such as through Cornish–Fisher expansions and in the calculation of second order moment corrections (bias and variance of estimators and Butler corrections for example). They are also the basis for the analysis of other forms of second-order inferences such as saddle point approximations and re-sampling

The authors are grateful to two referees for their insightful comments. Any errors are the responsibility of the authors.

✉ Gubhinder Kundhi
gkundhi@mun.ca

Paul Rilstone
pril@yorku.ca

1 Memorial University of Newfoundland, AA3082, 230 Elizabeth Avenue, St. John's, NL A1C 5S7, Canada

2 York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada

Published online: 20 October 2019

techniques. In the former case the saddle point is often derived as a tilted Edgeworth. In the latter case, an Edgeworth expansion is typically used to show the validity of jackknife and bootstrap methods.

Edgeworth expansions (valid or formal) can be difficult to derive except in the simplest univariate cases. The corresponding multivariate derivations such as in in Bhattacharya and Ghosh (1978) and Hall (1992) are quite labourous and may be difficult to follow. This is largely due to the complications of dimensionality. Since the expansions require higher order series approximations of intrinsically nonlinear and multivariate complex valued functions, the typical result is an abundance of notation using varying combinations of scalar results, multi-index and tensor notation, often combined with Einstein's method of tensor summation/multiplication (see e.g. Schild and Synge 1969) interspersed with some matrix methods.[1] Without impugning these contributions by any means, it is nevertheless a fact that many researchers may find the multivariate derivations difficult to follow.

In this paper we demonstrate that all of the standard results for multivariate Edgeworth expansions can be derived concisely and transparently using matrix algebra. Although many of the components of this result have appeared in various papers, a rigourous derivation of a valid multivariate Edgeworth expansion of arbitrary order using matrix differentiation has not been published, to our knowledge. The results are completely analogous to the univariate case, with Kronecker product and differentiation substituting for standard scalar products and differentiation. The advantages of this become evident in the exposition. The objects which are standard to higher order univariate expansions, such as moments, cumulants, and Hermite polynomials all have their Kronecker method counterparts.[2] We build on some useful results in this regard and add a few techniques useful to multivariate calculus and statistics. The results are readily programmed in any matrix based computer language.

The discussion is structured as follows. In "Some Matrix Calculus" we state some established results from matrix calculus and contribute a number of results which appear new. In "Fourier Transforms and Hermite Polynomials", "Moments and Characteristic Functions" and "Cumulants and Cumulant Generating Functions" sections establish some results applying matrix calculus methods to characteristic functions (CFs) and cumulant generating functions (CGFs), Fourier transforms and Hermite polynomials. In "Edgeworth expansion" we derive a valid multivariate Edgeworth expansion using streamlined matrix algebra. In "Conclusion" section concludes.

Some Matrix Calculus

We use fairly standard notation throughout the paper. The k dimensional identity matrix is denoted I_k and its j'th column is denoted ι_j^k. When the dimensions of these are clear from the context, we denote these I and ι_j, respectively.

[1] Papers by Phillips and Park (1988), Hansen (2006), Rilstone et al. (1996), Kollo and von Rosen (1998) and Kundhi and Rilstone (2012, 2013), amongst others, have somewhat streamlined this by employing some limited matrix algebra.

[2] A formal matrix-based third-order Edgeworth is presented in Traat (1986), but the result is not generalized nor are the derivations given necessary to demonstrate it is a valid asymptotic expansion.

Throughout this section let A and B be $m \times n$ and $p \times q$ matrices, respectively, containing real or complex elements. The (right) Kronecker product is defined in the usual way: $A \otimes B = [A_{ij}B]$, an $mp \times nq$ matrix. For matrices $A_1, \ldots, A_J$, we define

$$\bigotimes_{j=1}^{J} A_i = A_1 \otimes A_2 \otimes \cdots \otimes A_J \tag{1}$$

which provides a convenient shorthand to define Kronecker "powers":

$$A^{\otimes J} \equiv \underbrace{A \otimes A \otimes \cdots \otimes A}_{J \text{ times}}, \tag{2}$$

an $m^J \times n^J$ matrix. In our applications often A is an $m \times 1$ vector, in which case $A^{\otimes J}$ is an $m^J \times 1$ vector and we then define $A^{\otimes 0} = 1$. We define the Vec operator: $\vec{A} \equiv \text{Vec}[A]$ as that which stacks the columns of A into an $mn \times 1$ vector. A' indicates the transpose of A. We define the norm $\|A\| = \left(\text{Vec}[A]'\text{Vec}[\bar{A}]\right)^{1/2}$, where $\bar{A}$ denote conjugate. We make extensive use of the commutation matrix which is defined as the $mn \times mn$ matrix, $K_{m,n}$, which satisfies $\vec{A} = K_{m,n}\overrightarrow{A'}$.

For reference below we group a number of the well-established properties of Kronecker products, the Vec operator and the commutation matrix as follows.

Proposition 1 (a) *Subject to conformability,* $(A \otimes B)(C \otimes B) = AC \otimes BD$; *(b)* $(A \otimes B) = K_{m,p}(B \otimes A)K_{q,n}$; *(c)* $K_{p,q}K_{q,p} = I_{pq}$; *(d)* $K_{k1} = K_{1k} = I_k$; *(e)* $K_{p,q}^2 = K_{p,q}$; *(f)* $K_{m,n} = K_{n,m}'$; *(g) Let* b *be* $p \times 1$ *and* $mn = p^J$, *then* $K_{m,n}b^{\otimes J} = b^{\otimes J}$; *(h)* $Vec[ABC] = (C' \otimes A)\vec{B}$; *(i) For* $j = 1, \ldots, J$: $(K_{k^j,k^{J-j+1}} \otimes I_k)K_{k^{J-j+1},k^{j+1}} = (I_{k^j} \otimes K_{k^{J-j+1},k})$.

Proof Balestra (1976) summarizes these results and/or simple variants.
For our purposes, commutation matrices are used largely to "rotate" Kronecker products of vectors, each of which is, say, $k \times 1$. Examples are in the "Appendix". □

Throughout this section we assume implicitly that the functions of interest are continuously differentiable up to the order specified. We follow the convention in MacRae (1974) and define the K-derivative[3] of a (real or complex valued) function $A(X)$ with respect to a real $k \times l$ matrix X as

$$\nabla_X\, A(X) = A \otimes \frac{\partial}{\partial X} = \left[\partial A_{ij}/\partial X\right] \tag{3}$$

such that $\nabla A(X)$ is an $mk \times nl$ matrix indexed in the same manner as Kronecker products.

Two of the main properties of K-derivatives which we use repeatedly are as follows.

[3] This designation is convenient and has been used by Jammalamadaka et al. (2006). We will similarly refer to K-moments and K-cumulants.

Proposition 2 *Let A and B be functions of the $k \times l$ matrix X with AB defined. Then (a) $\bigtriangledown AB = (A \otimes I_k) \bigtriangledown B + \bigtriangledown A(B \otimes I_l)$ and (b) $\bigtriangledown(A \otimes B) = A \otimes \bigtriangledown B + (K_{m,p} \otimes I_k)(B \otimes \bigtriangledown A)(K_{q,n} \otimes I_l)$.*

Proof See MacRae (1974).
MacRae's proof is not explicitly for complex valued functions, but it applies in that case since the usual product rule of calculus applies to complex valued functions and each of the elements of AB or $A \otimes B$ can be seen as the product of differentiable functions. This same observation applies below. In most of our results, these results are used with A, B and X being vectors (or sometimes scalars), in which case the results simplify considerably. Note that if a is a scalar constant then $\bigtriangledown aB = a \bigtriangledown B$. However, when $a = a(X)$, is a function of X, it is prudent to write $aB = a \otimes B$ and use Proposition 2 (b) for differentiation. □

As will become clear, there are a number of advantages to using a K-derivative. One is that, as a result of its definition, it inherits the properties associated with the Kronecker product, notably the associativity property. It is often very useful to keep track of dimensions and order of differentiation. We will often suppress obvious notations so that $\bigtriangledown A = \bigtriangledown_X A(X)$.

Matrices of higher order derivatives are constructed recursively so that

$$\bigtriangledown^J A(X) = \bigtriangledown \bigtriangledown^{J-1} A(X). \tag{4}$$

The set of k^J derivatives of order J of a function form a tensor. There are different ways of cataloging these using different, but isometric, coordinate spaces. Multi-index tensor notation is used for example in McCullagh (1987). Another alternative, slightly different from K-differentiation is the vector derivative used for example in Magnus and Neudecker (1988). There the convention is to differentiate $\vec{A}$ with respect to $\overrightarrow{X'}$. Kollo and von Rosen (1998) have used vector differentiation to obtain third-order (i.e., skewness corrected) multivariate expansions. The method does not seem to easily extend to more general cases however and even the lower order results require a splicing of Kronecker and star products. All of these approaches are of course isometric since vector derivatives can be obtained using K-derivative notation through $\partial\vec{A}/\partial\overrightarrow{X'} = \bigtriangledown_{\overrightarrow{X'}}\vec{A}$. One advantage to the K-derivative is in the treatment of third and higher order derivatives and Taylor series. Also, since the generated mapping follows the same convention as Kronecker products, the K-derivatives maintain corresponding properties.

It is often convenient to differentiate with respect to the transpose of the argument of a function in which case we denote $A^{(J)}(X) = \bigtriangledown^J_{X'} A(X)$. We define $\bigtriangledown^0 A = A(X)$. We provide examples of matrix derivatives in the "Appendix".

A useful result in calculus is Young's Theorem pertaining to the order of differentiation. With respect to K-derivatives, this can be expressed in the following manner, which we have not seen elsewhere.

Proposition 3 (Young's Theorem) *Let $f(x)$ be a J-times continuously differentiable function $f(x)$ where x is $k \times 1$. Then for $j = 1, \ldots, J$.*

$$(I_{k^{j-1}} \otimes K_{k^{J-j},k}) \bigtriangledown^J f(x) = \bigtriangledown^J f(x).$$

Proof and examples: See the "Appendix".

Two useful extensions of Proposition 2 (b) are provided in the following results, which we have not seen elsewhere.

Proposition 4 *Let x be $k \times 1$. Then, (a)*

$$\bigtriangledown_x (x'^{\otimes J}) = J(x'^{\otimes J-1} \otimes I_k) K_{k,k^{J-1}}$$

and (b)

$$\bigtriangledown_x^J x'^{\otimes J} = J! \prod_{j=1}^{J} \left(I_{k^{j-1}} \otimes K_{k,k^{J-j}}\right).$$

Proof See the "Appendix". □

The Kronecker approach to matrix differentiation has been criticized[4] as not lending itself to a straightforward generalization of the chain rule.[5] Actually, this is not correct as we construct the following version of the chain rule for K-derivatives.

First we state what we mean by a matrix chain rule. Suppose we have $Z = Z(Y)$ and $Y = Y(X)$ where Y is $s \times t$. The chain rule for an arbitrary element of Z with respect to an arbitrary element of X can be written $\frac{\partial Z_{ij}}{\partial X_{uv}}$ where each Z_{ij} is a composite function of the st elements of Y and each of these depends on X_{kl}. By the usual chain rule applied to scalar functions we have

$$\frac{\partial Z_{ij}}{\partial X_{uv}} = \sum_{a=1}^{s} \sum_{b=1}^{t} \frac{\partial Z_{ij}}{\partial Y_{ab}} \frac{\partial Y_{ab}}{\partial X_{uv}}. \tag{5}$$

What we require is a streamlined version of this equation which applies for each Z_{ij} in Z and X_{uv} in X.

Proposition 5 (Chain Rule) *Let $Z = Z(Y)$ and $Y = Y(X)$ where X is $k \times l$, Z is $p \times q$ and Y is $s \times t$. Then $\bigtriangledown_X Z = (\bigtriangledown_{\vec{Y}'} Z \otimes I_k)(I_q \otimes \bigtriangledown_X \vec{Y})$.*

Proof See the "Appendix". □

[4] e.g., Kollo and von Rosen (2005, p. 121)

[5] MacRae (1974) provides a matrix version of the chain rule although it uses star products and is not readily integrated with other matrix operations. Magnus (2010) states the result in Proposition 5, but does not supply a proof.

This chain rule is used below in a number of instances, differentiating exponential and logarithmic functions, defining Hermite polynomials from general (not-necessarily standardized) distributions, showing homogeneity of cumulants and certain details in the derivation of the Edgeworth.

Propositions 3, 4 and 5 are useful, *inter alia*, in the concise construction and manipulation of multivariate Taylor series using higher-order Kronecker derivatives. In this regard we have the following.

Proposition 6 (Taylor series) *Let* $f : \mathbb{R}^k \rightarrow \mathbb{C}$ *be* J*-times continuously differentiable in a neighbourhood of* x_0*. Then, for some neighbourhood of* x_0*, there exists a* J*th order polynomial,* $f_J(x)$*, such that (a)*

$$f(x) = f_J(x) + r_J(x)$$

where

$$f_J(x) = \sum_{j=0}^{J} \frac{1}{j!} f^{(j)}(x_0)(x - x_0)^{\otimes j},$$

with remainder in integral form:

$$r_J(x) = \frac{1}{(J-1)!} \left(\int_0^1 \left[f^{(J)}(x_0 + t(x - x_0)) - f^{(J)}(x_0) \right] (1 - t)^{J-1} \, dt \right) (x - x_0)^{\otimes J},$$

or in Lagrange form:

$$r_J(x) = \frac{1}{J!} (f^{(J)}(\bar{x}^{\dagger}) - f^{(J)}(x_0))(x - x_0)^{\otimes J}$$

$x^{\dagger} = x_0 + c(x - x_0)$ *for some* $c \in [0, 1]$ *and (b) where*

$$f^{(j)}(x_0) = f_J^{(j)}(x_0), \qquad j = 0, 1, \ldots, J.$$

Proof See the "Appendix". □

Kollo and von Rosen (2005) have a representation for multivariate Taylor series using vector rather than Kronecker derivatives. Rilstone et al. (1996) constructed a third-order Taylor series using Kronecker derivatives and without a detailed treatment of the remainder term. We usually find Lagrange's form of the remainder more convenient to work with. We note that this is usually referred to as a $(J-1)$'th order Taylor series. In our case, f is usually the CF of a random variable and J will typically refer to the highest moment that we assume exists, three or four typically.

One of the results we derive in "Cumulants and Cumulant Generating Functions" is a direct representation of the cumulants of a random vector in terms of its non-central moments. The following double sequence of matrices is used in their construction.

$$a_j^J = \begin{cases} I_{k^{J+1}} & j = 0;\ J = 1, 2, 3, \ldots \\ (a_j^{J-1} \otimes I_k) + (a_{j-1}^{J-1} \otimes I_k)(I_{k^{j-1}} \otimes K_{k^{J-j+1},k}) & j = 1, \ldots, J-1;\ J = 2, 3, \ldots. \\ I_{k^{J-1}} \otimes K_{k,k} & j = J \end{cases} \tag{6}$$

When $k = 1$, the a_j^J' s are simply the binomial coefficients from Pascal's triangle.

Proposition 7 *Let $A(x) = C(x)B(x)$ where x is $k \times 1$, C is a scalar function and B is $k \times 1$. Then*

$$\triangledown^J A = \sum_{j=0}^{J} a_j^J (\triangledown^j C \otimes \triangledown^{J-j} B).$$

Proof See the "Appendix". The proof essentially follows the steps of a traditional proof of the binomial theorem. □

The following application of the chain rule is in turn useful to derive, *inter alia*, a version of the homogeneity property for moments (in particular the multivariate normal) and cumulants and also multivariate Hermite polynomials based on a normal with full covariance matrix.

Proposition 8 *Let $f(t) = g(s(t))$ where t is $k \times 1$, $s = Bt$, B is $n \times k$ and g is J times differentiable. Then*

$$\triangledown_t^J f(t) = B^{\otimes J}\, \triangledown_s^J\, g(s(t)).$$

Proof See the "Appendix". □

A common step in the construction of Edgeworth expansions is the inversion of the approximate CF which is typically facilitated by the link, say, between moments of CFs (themselves Fourier functions) and derivatives of the corresponding probability density function (PDF). A result which is useful in this regard is the following matrix version of integration by parts.

Proposition 9 (Integration by parts) *Let X be $k \times l$, $Y(X)$ be $s \times t$ and $Z(X)$ be $p \times q$ and suppose $Y(X) \otimes Z(X)$ disappears at its boundary. Then,*

$$\int Y \otimes \triangledown Z\, dX = -(K_{s,p} \otimes I_k) \int (Z \otimes \triangledown Y)\, dX\, (K_{q,t} \otimes I_l).$$

Proof See the "Appendix". □

We have not seen Propositions 7, 8 or 9 elsewhere although Kollo and von Rosen (2005) have a variant of Proposition 9 using vector derivatives.

Fourier Transforms and Hermite Polynomials

Fourier transforms and Hermite polynomials play central roles in the derivation of Edgeworth expansions. Their basic properties are well-established. Here we focus on representing a number of these properties using the results of "Some Matrix Calculus".

The Fourier transform of a function $f : \mathbb{R}^k \to \mathbb{C}$ is given by

$$\mathcal{C}_f(t) = \int f(x)e^{it'x}dx. \tag{7}$$

One of the familiar properties of Fourier transforms is the relationship between them and the Fourier transform of the derivatives of the original function. In this regard we establish the following matrix version of this relationship which is easily verified using Proposition 9.

Proposition 10 *Assume f is J-times differentiable and has Fourier transform $\mathcal{C}_f$. Then $\bigtriangledown^J f(x)$ has Fourier transform:*

$$\mathcal{C}_{\bigtriangledown^J f}(t) = (-it)^{\otimes J}\mathcal{C}_f(t).$$

Proof See the "Appendix". □

We note that $\mathcal{C}_{\bigtriangledown^J f}(t)$ is $k^J \times 1$. In the rest of the paper, f will be the density function of a $k \times 1$ random vector, X, and $\mathcal{C}$ is a characteristic function. In this case we find it convenient to use the notation $\mathcal{C}(t; X) = \int f(x)e^{it'x}dx$.

Hermite polynomials play a useful role in an Edgeworth expansion as they arise as the inverses to the polynomials occurring in an expansion of the corresponding CFs. They provide for a concise representation of terms in the expansions and their properties allow for convenient simplifications in inversion of CFs. The construction of matrix-valued Hermite polynomials has been considered by several authors. A number of useful results are found in Holmquist (1996). Here we will use a version of these in rigorously deriving a multivariate Edgeworth expansion.[6] Bhattacharya and Ghosh (1978) and Hall (1992) for example use generic polynomials, which some may find cumbersome. We define Hermite polynomials on $\mathbb{R}^k$ in a manner analogous to the standard univariate case. Let $\phi(\cdot; \Sigma)$ denote the k-variate normal density $N(0, \Sigma)$ of a random variable with zero mean and covariance matrix Σ. and put $\phi(\cdot) = \phi(\cdot; I_k)$. We simplify Holmquist's (1996) approach by defining Hermite polynomials with respect to the $N(0, I_k)$ and then use Proposition 8 to transform these to polynomials with respect to the $N(0, \Sigma)$. Define

$$\phi(z)H_n(z) = (-1)^n \bigtriangledown^n \phi(z), \quad n = 0, 1, 2, \ldots. \tag{8}$$

In the univariate case, the $H_n(z)$'s are scalars and the $H_n(z)$'s can be constructed recursively. Here $H_n : \mathbb{R}^k \to \mathbb{R}^{kn}$, and the sequence can also be built recursively as follows. Differentiating both sides of the definition:

[6] Chambers (1967) derives a multivariate Edgeworth, but defines Hermite polynomials using multi-index notation. Kollo and von Rosen (1998) use the third Hermite in matrix form.

$$(-1)^n \bigtriangledown^{n+1} \phi(z) = \phi(z) \otimes \bigtriangledown H_n(z) - \phi(z)(H_n(z) \otimes z) \tag{9}$$

so that

$$H_{n+1}(z) = H_n(z) \otimes z - \bigtriangledown H_n(z). \tag{10}$$

Most situations use up to the sixth Hermite. Using rotation matrices, $K[J,k] = \sum_{l=1}^{J} K_{k^l k^{J-l}}$, we can provide explicit expressions for these are as follows.

$$\begin{aligned}
H_0(z) &= 1 \\
H_1(z) &= z \\
H_2(z) &= z^{\otimes 2} - \vec{I}_k \\
H_3(z) &= z^{\otimes 3} + C_{31}(\vec{I}_k \otimes z) \\
&\quad C_{31} = -I_{k^3} - (K[2,k] \otimes I_k)K_{k,k^2} \\
H_4(z) &= z^{\otimes 4} + C_{42}(\vec{I}_k \otimes z^{\otimes 2}) + C_{40}\vec{I}_k^{\otimes 2} \\
&\quad C_{42} = C_{31} \otimes I_k - (K[3,k] \otimes I_k)K_{k^2,k^2} \\
&\quad C_{40} = -C_{31} \otimes I_k \\
H_5(z) &= z^{\otimes 5} + C_{53}(\vec{I}_k \otimes z^{\otimes 3}) + C_{51}(\vec{I}_k^{\otimes 2} \otimes z) \\
&\quad C_{53} = C_{42} \otimes I_k - (K[4,k] \otimes I_k)K_{k^3,k^2} \\
&\quad C_{51} = C_{40} \otimes I_k - (C_{42} \otimes I_k)(I_{k^2} \otimes [(K[2,k] \otimes I_k)K_{k,k^2}]) \\
H_6(z) &= z^{\otimes 6} + C_{64}(\vec{I}_k \otimes z^{\otimes 4}) + C_{62}(\vec{I}_k^{\otimes 2} \otimes z^{\otimes 2}) + C_{60}\vec{I}_k^{\otimes 3} \\
&\quad C_{64} = C_{53} \otimes I_k - (K[5,k] \otimes I_k)K_{k^4,k^2} \\
&\quad C_{62} = C_{51} \otimes I_k - (C_{53} \otimes I_k)(I_{k^2} \otimes [(K[3,k] \otimes I_k)K_{k^2 k^2}]) \\
&\quad C_{60} = -C_{51} \otimes I_k
\end{aligned} \tag{11}$$

It is straightforward to confirm that when $k = 1$ the polynomials simplify to:

$$\begin{aligned}
H_1(z) &= z \\
H_2(z) &= z^2 - 1 \\
H_3(z) &= z^3 - 3z \\
H_4(z) &= z^4 - 6z^2 + 3 \\
H_5(z) &= z^5 - 10z^3 + 15z \\
H_6(z) &= z^6 - 15z^4 + 45z^2 - 15
\end{aligned} \tag{12}$$

A feature of univariate Hermite polynomials is their orthogonality. Although this property is not used in the construction of the Edgeworth expansions in this paper, it is interesting to verify this holds in the multivariate case here. Consider the situations for $n \geq l \geq 0$. (Otherwise switch the roles of H_n and H_l.) By definition and integration by parts (Proposition 7) we have

$$\int H_l(z)H_n(z)'\phi(z)\,dz = (-1)^n \int H_l(z)\phi^{(n)}(z)\,dz$$
$$= (-1)^n(-1)\int H_l(z)^{(1)}(\phi^{(n-1)}(z)\otimes I_k)\,dz. \tag{13}$$

Note that H_l is a polynomial of order l so that if $n > l$, $H_l(z)^{(n)} = \nabla^n_{z'} H_l(z) = 0_{k^l,k^n}$. If $n = l$, $H_l(z)^{(n)} = \nabla^n_{z'} H_l(z) = \nabla^n_{z'} z^{\otimes n} = n!\prod_{j=1}^n \left(I_{k^{j-1}} \otimes K_{k,k^{n-j}}\right)$.

$$\int H_l(z)H_n(z)'\phi(z)\,dz = (-1)^n(-1)^n \int H_n(z)^{(n)}(\phi^{(0)}(z)\otimes I_k^n)\,dz$$
$$= \int n!\prod_{j=1}^n \left(I_{k^{j-1}} \otimes K_{k,k^{n-j}}\right)(\phi(z)\otimes I_k^n)\,dz$$
$$= n!\prod_{j=1}^n \left(I_{k^{j-1}} \otimes K_{k,k^{n-j}}\right) \tag{14}$$

and hence

$$\int H_l(z)H_n(z)'\phi(z)\,dz = \begin{cases} 0_{k^l\times k^n}, & l \neq n \\ n!\prod_{j=1}^n \left(I_{k^{j-1}} \otimes K_{k,k^{n-j}}\right), & l = n \end{cases}. \tag{15}$$

The above results apply for multivariate Hermites which are generated from a $N(0, I_k)$ distribution. To generate Hermites from a $N(0, \Sigma)$ distribution, it is straightforward to do so by defining

$$\phi(x;\Sigma)H_n(x;\Sigma) = (-1)^n \,\nabla^n\, \phi(x;\Sigma), \quad n = 0, 1, 2, \ldots \tag{16}$$

setting $B = \Sigma^{-1/2}$ and using Proposition 8 so that $\nabla^n_x \phi(x;\Sigma) = B^{\otimes n}\nabla^n_s \phi(s; I_k)|B|^k$.

Moments and Characteristic Functions

For a $k \times 1$ random vector X we define the J'th (non-central) "Kronecker" or K-moment as the $k^J \times 1$ vector

$$\mu_J = E\left[X^{\otimes J}\right]. \tag{17}$$

There are number of reasons for arranging moments in this way. One is that, as will become apparent, they are (times i^J) K-derivatives of the CF of X or equivalently, μ_J is simply the coefficient in the J'th term of the usual series expansion of the CF.

Although we do not make explicit use of them here, note that one can retrieve individual cross moments, say $E[X_{j_1}X_{j_2}\cdots X_{j_J}]$, by defining $k \times 1$ basis vectors ι_{j_1} and the k^J basis vector

$$\iota_{j_1 j_2\cdots j_J} = \iota_{j_1}\otimes\iota_{j_2}\otimes\cdots\otimes\iota_{j_J} = \bigotimes_{l=1}^{J}\iota_{j_l} \tag{18}$$

with

$$E[X_{j_1}X_{j_2}\cdots X_{j_J}] = \iota'_{j_1 j_2 \cdots j_J}\mu_J = E\left[\prod_{l=1}^{J}\iota'_{j_l}X\right]. \tag{19}$$

It is often useful to make assumptions with respect to the J'th absolute moment of a random vector. This is defined as

$$\rho_s = E\left[\left(\sum_{i=1}^{k}X_i^2\right)^{s/2}\right]. \tag{20}$$

Note that repeatedly using Proposition 1(a) we have

$$X'^{\otimes s}X^{\otimes s} = (X'X)^s \tag{21}$$

so that we can conveniently write

$$\rho_s = E\left[\left(X'^{\otimes s}X^{\otimes s}\right)^{1/2}\right] = E\left[\|X^{\otimes s}\|\right]. \tag{22}$$

Since

$$\bigtriangledown^J \mathcal{C}(t;X) = E\left[(iX)^{\otimes J}e^{it'X}\right] \tag{23}$$

we see that the J'th K-moment can be derived as

$$\mu_J = i^{-J}\bigtriangledown^J \mathcal{C}(0;X). \tag{24}$$

The variance matrix of the random variable X, $V = E[(X-\mu_1)(X-\mu_1)']$ can be computed as: $i^2V = \bigtriangledown_{t'}\bigtriangledown_t \mathcal{C}(0;X) - \bigtriangledown_t\mathcal{C}(0;X)\bigtriangledown_t \mathcal{C}(0;X)'$. V is often more convenient than, say, $\vec{V} = \mu_2 - \mu_1^{\otimes 2}$; for example it often immediately allows us to identify some quantities as quadratic forms.

The Edgeworth expansions make use of CFs and CGFs. There are numerous standard results concerning CFs we use below and group here for convenience.

Proposition 11 *(a) Let X_i, $i = 1,\ldots,N$, be i.i.d. Then $\mathcal{C}(t;\sum X_i) = \mathcal{C}(t;X)^N$; (b) $\mathcal{C}(t;AX) = \mathcal{C}(At;X)$, A symmetric; (c) If $\rho_s(X) < \infty$, then $\mathcal{C}(t;X)$ is s-times continuously differentiable in a neighbourhood of zero.*

Proof Feller (1971) summarizes these results and/or simple variants. □

For formal results it is useful to make use of the series representation of the CF.

$$\mathcal{C}(t;X) = 1 + \sum_{j\geq 1}\frac{1}{j!}(it)'^{\otimes j}\mu_j. \tag{25}$$

Cumulants and Cumulant Generating Functions

The CGF of the $k \times 1$ random vector X is $\mathcal{K}(t; X) = \log \mathcal{C}(t; X)$. Multivariate cumulants are often defined in somewhat cumbersome ways which involve defining partitions and sub-partitions of the indexing set (i.e., $1, \ldots, k$). Subsequent manipulations over these constructions may be quite tedious and the results opaque. A simpler method, employed by Jammalamadaka et al. (2006) is to define the J'th cumulant using K-derivatives of the CGF. From our perspective this leads to a more direct intuitive definition, in a manner analogous to the J'th moment. In the univariate case, the cumulants are often defined formally by equating coefficients from series expansions of $\mathcal{K}(t; X)$ and $\log \mathcal{C}(t; X)$.

For formal results we have the series representation of the CGF.

$$\mathcal{K}(t; X) = \sum_{l \geq 1} \frac{1}{l!} (it)^{\prime \otimes l} \kappa_l. \tag{26}$$

Since $\mathcal{K}(t; X) = \log \mathcal{C}(t; X)$ one can formally define the Kronecker cumulants, by equating the coefficient on the $(it)^{\otimes J}$ terms in (26) to those on $(it)^{\otimes J}$ from the log of $\mathcal{C}$ in (25), viz,

$$\sum_{l \geq 1} \frac{1}{l!} (it)^{\prime \otimes l} \kappa_l = \sum_{k \geq 1} \frac{(-1)^{l+1}}{l} \left(\sum_{j \geq 1} \frac{1}{j!} (it)^{\prime \otimes j} \mu_j \right)^k. \tag{27}$$

This is analogous to the univariate case, apart from the fact that the moments and cumulants are vector-valued. This definition of cumulants is consistent with the definition of K-moments, in that κ_J and μ_J are both k^J in dimension. An additional advantage is that certain invariance properties are easily shown to hold for K-cumulants. We can mechanically derive explicit solutions for κ_J in a couple of ways. As in the univariate case, one method is to truncate the series for terms smaller than $O(\|t\|^s)$, $\|t\| \to 0$. In the univariate case $\kappa_1 = \mu_1$ and $\kappa_2 = \mu_2 - \mu_1^2$ and there is also a simple relationship between the higher cumulants and the central moments. We examine this for the multivariate case.

Alternatively (but equivalently, when the moments exist), we define the J'th K-cumulant, κ_J, when it exists, from $t^{‘\otimes J} \kappa_J = (i)^{-J} t^{‘\otimes J} \bigtriangledown^J \mathcal{K}(0; X)$. As with the J'th moment, κ_J is k^J in dimension. This has the advantage of not relying on formal representations of the CF and CGF. We derive an explicit formula for the cumulants. The following proposition regards the higher order derivatives of the logarithm of a function, $C(t)$. Here $C(t)$ is an arbitrary real or complex function although in our application of the proposition $C(t)$ will be a CF. Put $\triangle^j = \bigtriangledown^j C(t)/C(t)$. J'th order cumulant is a polynomial in the first J non-central moments. Put $J_s = J - s$ with $J_0 = J$.

Proposition 12 *Let $C(t)$ be J-times differentiable and bounded away from zero in a neighbourhood of t. Then:*

$$\triangledown^J \log C(t) = \triangle^J + \sum_{l=1}^{J-1}(-1)^l \sum_{j_1=1}^{J_0-1}\sum_{j_2=1}^{J_1-1}\cdots\sum_{j_l=1}^{J_l-1}\prod_{s=0}^{l-1}(I_{k^{J-J_s}} \otimes a_{j_{s+1}}^{J_s-1})\bigotimes_{s=1}^{l}\triangle^{j_s}\otimes\triangle^{J_l}.$$

Proof See the "Appendix". □

We include examples of this result for $J = 2, 3, 4$ in the "Appendix". As we now see, we can use Proposition 12 to expediently and explicitly evaluate the vector-valued cumulants of a vector valued random variable.

Proposition 13 (Cumulants) *Let the J'th moment of X exist. Then the J'th derivative of $\mathcal{K}(t; X)$ exists at zero and $t^{`\otimes J} \triangledown^J \mathcal{K}(0; X) = i^J t^{`\otimes J}\kappa_J$ where $\kappa_1 = \mu_1$ and, for $J = 2, 3, \ldots,$*

$$\kappa_J = \mu_J + \sum_{l=1}^{J-1}(-1)^l \sum_{j_1=1}^{J_0-1}\sum_{j_2=1}^{J_1-1}\cdots\sum_{j_l=1}^{J_l-1}\left(\prod_{s=0}^{l-1}(I_{k^{J-J_s}} \otimes a_{j_{i+1}}^{J_i-1})\right)\bigotimes_{s=1}^{l}\mu_{j_1}\otimes\mu_{J_l}.$$

Proof The result follows immediately from Proposition 12: at $t = 0$, $\mathcal{C}(0; X) = 1$, $\triangle^j = i^j\mu_j$ noting $\triangle^1 = \triangledown\mathcal{K}(t; X)$. □

As stated, Kronecker cumulants have been defined before, however we are unaware of their direct derivation as in Propositions 12 and 13. The generality of this expression may look somewhat forbidding; however it simplifies dramatically for those cases typically of interest. Evaluating this expression this for $J = 1, 2, 3, 4$, we have:

$$\begin{aligned}
\kappa_1 &= \mu_1\\
\kappa_2 &= \mu_2 - \mu_1^{\otimes 2} = \vec{V}\\
\kappa_3 &= \mu_3 - 3\mu_1\otimes\mu_2 + 2\mu_1^{\otimes 3}\\
\kappa_4 &= \mu_4 - 4\mu_1\otimes\mu_3 - 3\mu_2^{\otimes 2} + 12\mu_1^{\otimes 2}\otimes\mu_2 - 6\mu_1^{\otimes 4}.
\end{aligned} \tag{28}$$

It is convenient to exploit certain results that hold between the CFs and CGFs of random vectors which are linear (affine) transformations of the other. The following results are well-established, but some, in their Kronecker product form, may be less well-known. In particular the results pertaining to higher order cumulants allow for some straightforward manipulations of the asymptotic expansions. For example, some results are easily established for standardized random vectors (mean zero and identify covariance matrix) and then modified for the more general case using semi-invariance and invariance properties of cumulants.

Let X be an $k \times 1$ random vector with CGF and CFs: $\mathcal{K}(t; X)$ and $\mathcal{C}(t; X)$. Let $Z = a + X$ where a is a $k \times 1$ vector of constants. The usual shift variance property of the first cumulant and shift invariance property of higher order cumulants easily verified. i.e. $\kappa_1(Z) = a + \kappa_1(X)$ and $\kappa_j(Z) = \kappa_j(X)$, $j = 2, \ldots, J$, provided the moments exist. We do have a convenient representation of the homogeneity property in Kronecker product form:

Proposition 14 *The J'th derivative of the CF of $Z = BX$ is $\nabla_t^J \mathcal{C}(t; Z) = B^{\otimes J} \nabla_s^J \mathcal{C}(s; X)$ and the J'th non-central moment is given by $E[Z^{\otimes J}] = B^{\otimes J} E[X^{\otimes J}]$. Similarly, the J'th derivative of the CGF of $Z = BX$ is $\nabla_t^J \mathcal{K}(t; Z) = B^{\otimes J} \nabla_s^J \mathcal{K}(s; X)$ and the J'th cumulant is given by $\kappa_J(Z) = B^{\otimes J} \kappa_J(X)$.*

Proof Follows directly from Proposition 8. □

Edgeworth Expansions

In this section we show that a valid Edgeworth expansion can be simply derived using the matrix methods developed above. The basic structure follows that of Feller (1971) or Bhattacharya and Ghosh (1978). To our knowledge, a rigorous derivation of the multivariate expansion using these matrix methods has not appeared elsewhere. Let $X_i, i = 1, \ldots, N$ be i.i.d. draws on the $k \times 1$ continuous random vector X. Assume $E[X] = 0$ and put $V = E[XX']$. Let $\bar{X} = \frac{1}{N} \sum X_i$ and $T_N = \sqrt{N}\bar{X}$. We focus on the PDF of T_N although the results are easily modified to retrieve approximate moments of T_N or other functionals of the sampling distribution of T_N. To extend the results here to heterogeneous random vectors, the notation needs to be modified to replace (constant) cumulants by their averages across draws. We conjecture that a matrix-based approach may also be useful in deriving Edgeworth expansions for certain stationary dependent multivariate processes although we do not attempt that here.

From Proposition 11 we have $\mathcal{C}(t; T_N) = \mathcal{C}(t/\sqrt{N}; X)^N$. Since $\mathcal{C}$ is a complex function, $\mathcal{K} = \log \mathcal{C}$ is defined such that $\mathcal{C} = e^{\mathcal{K}}$ and $\mathcal{K}$ is the principle logarithm. We note that $|\mathcal{C}(t/\sqrt{N}; X) - 1|$ is interior to the unit circle so that $\mathcal{K}$ is well-defined. We thus have $\mathcal{K}(t; T_N) = N \log \mathcal{C}(t/\sqrt{N}; X) = N\mathcal{K}(t/\sqrt{N}; X)$. The results in the previous sections allow us to construct the following approximation to the CGF. Put

$$\begin{aligned}
\mathcal{K}_s(t; T_N) &= \sum_{j=2}^{s} \frac{(it)'^{\otimes j} \kappa_j}{j! N^{(j-2)/2}} \\
r_s(t) &= R_s(t) \frac{(iV^{1/2}t)^{\otimes s}}{s! N^{(s-2)/2}} \\
R_s(t) &= (V^{-1/2 \otimes s}) \left(\mathcal{K}^{(s)}(c/\sqrt{N}; X) - \mathcal{K}^{(s)}(0; X) \right)
\end{aligned} \tag{29}$$

where c is between zero and 1. Here and henceforth, an s subscript on an approximating function indicates it is accurate up to $O(N^{-(s-2)/2})$.

Proposition 15 *Let X_i be a continuous random vector having finite s'th absolute moments, some $s \geq 2$ and non-singular covariance matrix V. Then, for some region around 0: $\|V^{1/2}t\| \leq \delta$, $\delta > 0$*

$$\mathcal{K}(t; T_N) = \mathcal{K}_s(t; T_N) + r_s(t).$$

Proof See the "Appendix". □

In the derivations in the "Appendix" it will be useful to decompose

$$\mathcal{K}_s(t; T_N) = -\frac{1}{2}t'Vt + \mathcal{K}_s^{\dagger}$$
$$\mathcal{K}_s^{\dagger} = \sum_{j=3}^{s} \frac{(it)^{'\otimes j}\kappa_j}{j!N^{(j-2)/2}}. \tag{30}$$

Terms such as $\mathcal{K}_s^{\dagger}$ are central to second order asymptotics as they provide measures of departures from normality. The technical details of a valid Edgeworth expansion consist largely in showing that $\mathcal{K}_s^{\dagger}$ is dominated by $-\frac{1}{2}t'Vt$, and in assessing the contribution of $\mathcal{K}_s^{\dagger}$ to the inversion of the CF of T_N.

We could construct an explicit s'th order polynomial approximation in t to $\mathcal{C}(t; T_N)$ an s'th order Taylor series expansion of $e^{\mathcal{K}^{\dagger}}$ at $\mathcal{K}^{\dagger} = 0$. This has the approximation properties we wish except that it contains terms of order smaller than $N^{-(s-2)/2}$. Of course these can simply be ignored (Feller does so in the univariate case) although we prefer an explicit definition of the approximate CF. A systematic way to eliminate the smaller order terms is as follows. Put $s_j = s - l_1 - l_2 - \cdots - l_{j-1}$ with $s_1 = s$ and

$$P_s(t) = 1 + \sum_{j=1}^{s} \frac{1}{j!} \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_j=1}^{s_j} \frac{\left(\bigotimes_{k=1}^{j} \kappa_{l_k+2}\right)'}{\prod_{k=1}^{j} N^{l_k/2}(l_k+2)!} (it)^{\otimes \sum_{k=1}^{j}(l_k+2)} \tag{31}$$

and define

$$\mathcal{C}_s(t; T_N) = e^{-\frac{1}{2}t'Vt} P_s(t) \tag{32}$$

such that $P_s(t)$ will agree with an s'th order Taylor series expansion of $e^{\mathcal{K}^{\dagger}}$ with $O(N^{-(s+1)/2})$ terms removed. The expression defining $P_s(t)$ may appear lengthy, but typically there are only a few terms in this polynomial when, e.g., $s = 3$ or $s = 4$. In the former case

$$P_3(t) = 1 + \frac{1}{N^{1/2}} \frac{\kappa_3'(it)^{\otimes 3}}{6} \tag{33}$$

and in the latter case

$$P_4(t) = P_3(t) + \frac{1}{N}\frac{\kappa_4'(it)^{\otimes 4}}{24} + \frac{1}{N}\frac{\kappa_3^{'\otimes 2}(it)^{\otimes 6}}{72}. \tag{34}$$

In verifying the second result it is useful to note that $\left(\kappa_3'\left(t^{\otimes 3}\right)\right)^2 = \kappa_3^{'\otimes 2} t^{\otimes 6}$.

Denote the PDF of T_N by $f(x; T_N)$. The Edgeworth approximation to $f(x; T_N)$, denoted $f_s(x; T_N)$, is the Fourier inverse of $\mathcal{C}_s(t; T_N)$. To quantify the approximation error we evaluate $(2\pi)^{-k}\int \triangle(t)e^{-it'x}dt$ where

$$\triangle(t) = \mathcal{C}(t;T_N) - \mathcal{C}_s(t;T_N). \tag{35}$$

As is standard, we break up the domain of integration into "small" and "large" values of t. For any $\delta \geq 0$, put $\mathcal{B}_N(\delta) = \{t : \|V^{1/2}t\| \leq \delta\sqrt{N}\}$. We have the following for "large" t.

Proposition 16 *Let the assumptions of Proposition 15 hold. Then, for any* $\delta > 0$, $s \geq 2$, $\left|\int_{\mathcal{B}_N^c(\delta)} \triangle(t)e^{-it'x}dt\right| = o\left(N^{-(s-2)/2}\right)$ *uniformly in* x.

Proof See the "Appendix". □

For "small" t we further decompose

$$\triangle(t) = \triangle_1(t) + \triangle_2(t), \tag{36}$$

$$\triangle_1(t) = (e^{\mathcal{K}(t;T_N)-\mathcal{K}_s(t;T_N)} - 1)e^{\mathcal{K}_s(t;T_N)}, \qquad \triangle_2(t) = e^{\mathcal{K}_s(t;T_N)} - \mathcal{C}_s(t;T_N) \tag{37}$$

Proposition 17 *Let the assumptions of Proposition 15 hold. Then, for some* $\delta > 0$, $s \geq 2$, *(a)* $\left|\int_{\mathcal{B}_N(\delta)} \triangle_1(t)e^{-itx}dt\right| = o\left(N^{-(s-2)/2}\right)$ *and (b)* $\left|\int_{\mathcal{B}_N(\delta)} \triangle_2(t)e^{-itx}dt\right| = o\left(N^{-(s-2)/2}\right)$ *uniformly in* x. □

Proof See the "Appendix". □

To obtain an explicit PDF approximation and the order of the approximation error, put

$$f_s(x;T_N) = \phi(x;V) + \sum_{j=1}^{s}\frac{1}{j!}\sum_{l_1=1}^{s_1}\sum_{l_2=1}^{s_2}\cdots\sum_{l_j=1}^{s_j}\frac{\left(\bigotimes_{k=1}^{j}\kappa_{l_k+2}\right)}{\prod_{k=1}^{j}N^{l_k/2}(l_k+2)!}H_{\sum_{k=1}^{j}(l_k+2)}(x;V)\phi(x;V). \tag{38}$$

Theorem 18 (Edgeworth Density): Let the assumptions of Proposition 15 hold. Then

$$\sup_x |f(x;T_N) - f_s(x;T_N)| = o\left(N^{-s/2}\right).$$

Proof This follows immediately from Propositions 16 and 17. □

As discussed above, what we consider innovative in this result is the fairly seamless and rigorous development of the basic multivariate Edgeworth expansion, analogous to the univariate case, using the tools of matrix calculus.

For the typical cases of $s = 3$ and $s = 4$ we have

$$f_3(x) = \phi(x; V) + \frac{1}{N^{1/2}} \frac{\kappa_3'}{6} H_3(x; V)\phi(x; V) \tag{39}$$

$$f_4(x) = f_3(x) + \frac{1}{N} \left(\frac{\kappa_4'}{24} H_4(x; V) + \frac{\kappa_3'^{\otimes 2}}{72} H_6(x; V) \right) \phi(x; V). \tag{40}$$

We emphasize here the compactness of the Edgeworth expansions written in this form. The basic components include the cumulants and the Hermite polynomials. Without using matrix algebra, both of these inputs can be quite tedious to derive, not to mention write down explicitly. For example the κ_j terms in (39) and (40) are replaced by, say $\kappa^{i_1,\ldots,i_j} = (\partial^j \mathcal{K}(t, X)/\partial t_{i_1} \cdots \partial t_{i_j})_{t=0}$ and the H_j terms are replaced by, say $H_{i_1 \cdots i_j}$, the polynomials obtained by differentiating $\phi(z, V)$ distribution with respect to $z_{i_1}, \ldots, z_{i_j}$.[7] This becomes very cumbersome using standard calculus. The inner products in (3.11) and (3.12) are then obtained by summing over all index sets. This is not to mention that the intermediate derivations we have done in the previous sections are extremely tedious using multi-index notation. We stress again that the gain using matrix algebra is one of convenience and not substance.

As a further example of the potential advantages of the matrix methods used here, consider approximating a conditional PDF using Edgeworth expansions. There are alternative, equivalent, ways to this. Partition $X' = \begin{pmatrix} X_1' & X_2' \end{pmatrix}$ where X_j is $k_j \times 1$, $j = 1, 2$ and suppose we want the approximation to $f(x_2|x_1) = f(x)/f(x_1)$ with obvious notation. We could obtain $f(x_1)$ either by integrating x_2 out of $f(x)$ or we could simply start afresh by recalculating the corresponding cumulants for X_1 and the corresponding Hermites $H_j : \mathbb{R}^{k_1} \to \mathbb{R}^{k_1 j}$. However we also see that, defining $S = \begin{pmatrix} I_{k_1} & O \end{pmatrix}$ where O is a $k_1 \times k_2$ matrix of zeros we have $E[X_1^j] = S^{\otimes j} \mu_J$, the cumulants for X_1 are $\kappa_{1,J} = S^{\otimes J} \kappa_J$ and the Hermites corresponding to a $N(0, I_{k_1})$ are $H_{1J}(z_1) = S^{\otimes j} H_J(z)$ so that no additional computations of any substance are required to obtain the marginal counterparts to the approximations in Eq. 38. Edgeworth approximations to $f(x_2|x_1)$ can be based on the ratio of the approximations to $f(x)$ and $f(x_1)$ in Eq. 38, with the caveat that neither numerator nor denominator may be strictly positive.

Conclusion

Let us contrast the derivations done here using matrix calculus with established, mostly scalar based, derivations. The matrix algebra-basic mathematical results allows for simultaneous manipulation of many quantities. Multivariate Taylor series are more parsimoniously written with matrix algebra. Similarly, a matrix version of the chain rule applied to the CGF allows us to parsimoniously define vector-valued cumulants without the need for multi-index notation and messy partitions over index sets. Similarly, vector-valued Hermite polynomials are naturally defined and of the same dimension as their corresponding cumulants, leading to the simple expressions in Eqs. 3.11 and 3.12.

[7] E.g. see e.g. Barndorff-Nielsen and Cox (1989, p.174).

This permits a very succinct way to catalog cumulants as vectors. Cumulants, central to Edgeworth type expansions are very tedious to define using multi-index notation. Here the cumulants have the identical form as the univariate case. The technical steps in the construction of the Edgeworth cannot be avoided using matrix algebra, but the arguments are simpler in that they avoid manipulation of the multiple sums involved using multi-index notation.

The basic Edgeworth expansion for a sample average is often used as a building block to examine the sampling properties of more complex estimators, from least squares estimators to implicitly defined extremum estimators and test statistics. Most of these can be written, or approximated, to higher order, by simple functions of sample averages. It is often convenient to write these using matrix methods. The results here can be used and extended to simplify these derivations. The assumptions we made can be readily relaxed in numerous ways. For example, the assumption that the random vector of interest is continuous can be simply modified so long as a Cramer-like condition is satisfied. Also, the analysis can be used to examine random matrices, simply by vectorizing them. Although we have considered the i.i.d. case here, the results are straightforward to extend to independent, but heterogeneous observations with appropriate redefining of the cumulants. The matrix-based results here could also simplify the derivation of Edgeworth expansions for certain time series problems [8] although the classic factorization of a joint CF for i.i.d. observations (Proposition 11 (a)) does not hold. Matrix-based results should also allow for more simple analysis of other tools used in higher-order asymptotic methods such as the saddle point and re-sampling techniques.

Appendix

This Appendix provides proofs of various propositions in the paper and provides numerous examples of the quantities used in the paper.

As examples of commutation matrices with $k = 2$ so that $K_{k^2,k} = K_{4,2}$ and $K_{k,k^2} = K_{2,4}$, we have

$$K_{4,2} = \begin{pmatrix} \begin{pmatrix} 1&0&0&0 \\ 0&0&0&0 \end{pmatrix} & \begin{pmatrix} 0&0&0&0 \\ 1&0&0&0 \end{pmatrix} \\ \begin{pmatrix} 0&1&0&0 \\ 0&0&0&0 \end{pmatrix} & \begin{pmatrix} 0&0&0&0 \\ 0&1&0&0 \end{pmatrix} \\ \begin{pmatrix} 0&0&1&0 \\ 0&0&0&0 \end{pmatrix} & \begin{pmatrix} 0&0&0&0 \\ 0&0&1&0 \end{pmatrix} \\ \begin{pmatrix} 0&0&0&1 \\ 0&0&0&0 \end{pmatrix} & \begin{pmatrix} 0&0&0&0 \\ 0&0&0&1 \end{pmatrix} \end{pmatrix}, \; K_{2,4} = \begin{pmatrix} \begin{pmatrix} 1&0 \\ 0&0 \\ 0&0 \\ 0&0 \end{pmatrix} & \begin{pmatrix} 0&0 \\ 1&0 \\ 0&0 \\ 0&0 \end{pmatrix} & \begin{pmatrix} 0&0 \\ 0&0 \\ 1&0 \\ 0&0 \end{pmatrix} & \begin{pmatrix} 0&0 \\ 0&0 \\ 0&0 \\ 1&0 \end{pmatrix} \\ \begin{pmatrix} 0&1 \\ 0&0 \\ 0&0 \\ 0&0 \end{pmatrix} & \begin{pmatrix} 0&0 \\ 0&1 \\ 0&0 \\ 0&0 \end{pmatrix} & \begin{pmatrix} 0&0 \\ 0&0 \\ 0&1 \\ 0&0 \end{pmatrix} & \begin{pmatrix} 0&0 \\ 0&0 \\ 0&0 \\ 0&1 \end{pmatrix} \end{pmatrix},$$

noting that, if x, y and z are $k \times 1$ vectors, we have $K_{k^2,k}(x \otimes y \otimes z) = y \otimes z \otimes x$ and $K_{k,k^2}(x \otimes y \otimes z) = z \otimes x \otimes y$.

[8] See, e.g., Taniguchi (1984).

As examples of higher-order Kronecker derivatives, consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. Here we indicate individual partial derivatives by superscripts:

$$f^{i_1 i_2 \cdots i_j}(x) = \frac{\partial^j f(x)}{\partial x_{i_j} \cdots \partial x_{i_2} \partial x_{i_1}}. \tag{A.1}$$

To illustrate the first four Kronecker derivatives (note that these are with respect to x') we have (suppressing the x argument f)

$$\begin{aligned}
f^{(1)} &= \nabla_{x'} f \\
&= \left(f^1 \; f^2 \right), \\
f^{(2)} &= \nabla_{x'} f^{(1)} \\
&= \left(f^{11} \; f^{12} \; f^{21} \; f^{22} \right), \\
f^{(3)} &= \nabla_{x'} f^{(2)} \\
&= \left(f^{111} \; f^{112} \; f^{121} \; f^{122} \; f^{211} \; f^{212} \; f^{221} \; f^{222} \right), \\
f^{(4)} &= \nabla_{x'} f^{(3)} \\
&= \begin{pmatrix} f^{1111} & f^{1112} & f^{1121} & f^{1122} & f^{1211} & f^{1212} & \cdots \\ \cdots & f^{1221} & f^{1222} & f^{2111} & f^{2112} & f^{2121} & \cdots \\ \cdots & f^{2122} & f^{2211} & f^{2212} & f^{2221} & f^{2222} & \end{pmatrix}.
\end{aligned} \tag{A.2}$$

Proof to Proposition 3 We can write the J'th K-derivative as

$$\nabla^J f(x) = \sum_{i_1=1}^{k} \cdots \sum_{i_J=1}^{k} \frac{\partial^J f(x)}{\partial x_{i_1} \cdots \partial x_{i_J}} \bigotimes_{l=1}^{J} \iota_{i_l}. \tag{A.3}$$

Thus,

$$\begin{aligned}
&(I_{k^j} \otimes K_{k^{J-j},k}) \nabla^J f(x) \\
&= \sum_{i_1=1}^{k} \cdots \sum_{i_J=1}^{k} \frac{\partial^J f(x)}{\partial x_{i_1} \cdots \partial x_{i_J}} (I_{k^j} \otimes K_{k^{J-j},m}) \left(\bigotimes_{l=1}^{j} \iota_{i_l} \bigotimes_{l=1}^{J-j} \iota_{i_l} \otimes \iota_{i_J} \right) \\
&= \sum_{i_1=1}^{k} \cdots \sum_{i_J=1}^{k} \frac{\partial^J f(x)}{\partial x_{i_1} \cdots \partial x_{i_J}} \bigotimes_{l=1}^{j} \iota_{i_l} \otimes \left(K_{k^{J-j},k} \left(\bigotimes_{l=1}^{J-j} \iota_{i_l} \otimes \iota_{i_J} \right) \right) \\
&= \sum_{i_1=1}^{k} \cdots \sum_{i_J=1}^{k} \frac{\partial^J f(x)}{\partial x_{i_1} \cdots \partial x_{i_J}} \left(\bigotimes_{l=1}^{j} \iota_{i_l} \right) \otimes \iota_{i_J} \otimes \left(\bigotimes_{l=1}^{J-j} \iota_{i_l} \right) \\
&= \nabla^J f(x),
\end{aligned} \tag{A.4}$$

the last equality arising from Young's Theorem. $\square$

To illustrate the matrix version of Young's Theorem stated in the paper, note that for $k = 2$ and second-order derivatives we need to show $(I_{2^{j-1}} \otimes K_{2^{2-j},2}) \bigtriangledown^2 f(x) = \bigtriangledown^2 f(x)$ for $j = 1, 2$. For $j = 2$ this is simply $(I_2 \otimes K_{2^0,2}) \bigtriangledown^2 f(x) = \bigtriangledown^2 f(x)$. For $j = 1$ we have

$$K_{k,k} \bigtriangledown f = K_{k,k}\mathrm{Vec}\begin{bmatrix} f^{11} & f^{12} \\ f^{21} & f^{22} \end{bmatrix} = \mathrm{Vec}\begin{bmatrix} f^{11} & f^{21} \\ f^{12} & f^{22} \end{bmatrix} = \bigtriangledown f. \tag{A.5}$$

For third-order derivatives we illustrate by evaluating $(I_{2^{j-1}} \otimes K_{2^{3-j},2}) \bigtriangledown^3 f(x)$ for $j = 1, 2, 3$. For $j = 3$ this is immediate. For $j = 2$

$$\begin{aligned}
(I_{2^{j-1}} \otimes K_{2^{3-j},2}) \bigtriangledown^3 f &= (I_2 \otimes K_{2,2}) \bigtriangledown^3 f \\
&= \begin{pmatrix} K_{2,2} & 0 \\ 0 & K_{2,2} \end{pmatrix} \bigtriangledown^3 f \\
&= \begin{pmatrix} K_{2,2} \begin{pmatrix} f^{111} \\ f^{112} \\ f^{121} \\ f^{122} \end{pmatrix} \\ K_{2,2} \begin{pmatrix} f^{211} \\ f^{212} \\ f^{221} \\ f^{222} \end{pmatrix} \end{pmatrix} \\
&= \begin{pmatrix} K_{2,2}\mathrm{Vec}\begin{bmatrix} f^{111} & f^{112} \\ f^{121} & f^{122} \end{bmatrix} \\ K_{2,2}\mathrm{Vec}\begin{bmatrix} f^{211} & f^{212} \\ f^{221} & f^{222} \end{bmatrix} \end{pmatrix} \\
&= \begin{pmatrix} \mathrm{Vec}\begin{bmatrix} f^{111} & f^{122} \\ f^{121} & f^{112} \end{bmatrix} \\ \mathrm{Vec}\begin{bmatrix} f^{211} & f^{221} \\ f^{221} & f^{212} \end{bmatrix} \end{pmatrix} \\
&= \bigtriangledown^3 f.
\end{aligned} \tag{A.6}$$

Similarly the result holds when $j = 1$. With respect to fourth-order derivatives ($J = 4$) the proposition states that $(I_{2^{j-1}} \otimes K_{2^{4-j},2}) \bigtriangledown^4 f = \bigtriangledown^4 f$ for $j = 1, 2, 3, 4$. For $j = 4$ this is immediate. Consider when $j = 1$. We have

$$
\begin{aligned}
(I_{2^{j-1}} \otimes K_{2^{4-j},2}) \bigtriangledown^4 f &= K_{8,2} \mathrm{Vec} \begin{bmatrix} f^{1111} & f^{2111} \\ f^{1112} & f^{2112} \\ f^{1121} & f^{2121} \\ f^{1122} & f^{2122} \\ f^{1211} & f^{2211} \\ f^{1212} & f^{2212} \\ f^{1221} & f^{2221} \\ f^{1222} & f^{2222} \end{bmatrix} \\
&= \mathrm{Vec} \begin{bmatrix} f^{1111} & f^{1112} & f^{1121} & f^{1122} & f^{1211} & f^{1212} & f^{1221} & f^{1222} \\ f^{2111} & f^{2112} & f^{2121} & f^{2122} & f^{2211} & f^{2212} & f^{2221} & f^{2222} \end{bmatrix} \\
&= \bigtriangledown^4 f. \qquad (A.7)
\end{aligned}
$$

Similarly the result also holds for $j = 2, 3$.

Proof to Proposition 4 Let $y_i = x'$ so that $x'^{\otimes J} = \bigotimes_{i=1}^{J} y_i$. Note that

$$
\begin{aligned}
\bigotimes_{i=1}^{J} y_i &= \left(\bigotimes_{i=1}^{J-1} y_i \right) \otimes y_J \\
&= \left(y_J \otimes \left(\bigotimes_{i=1}^{J-2} y_i \right) \otimes y_{J-1} \right) K_{k,k^{J-1}} \\
&= \left(\left(\bigotimes_{i=l+1}^{J} y_i \right) \left(\bigotimes_{i=1}^{l-1} y_i \right) \otimes y_l \right) K_{k,k^{J-1}} \qquad (A.8)
\end{aligned}
$$

for any integer l, $1 \le l \le J$ where we repeatedly use $K_{k,k^{J-1}} K_{k,k^{J-1}} = K_{k,k^{J-1}}$. So,

$$
\begin{aligned}
\bigtriangledown_x \left(\bigotimes_{i=1}^{J} y_i \right) &= \sum_{l=1}^{J} \bigtriangledown_x \left(\left(\bigotimes_{i=l+1}^{J} y_i \right) \left(\bigotimes_{i=1}^{l-1} y_i \right) \otimes y_l \right)_{y_i \text{ const } i \neq l} K_{k,k^{J-1}} \\
&= \sum_{l=1}^{J} \left(\left(\bigotimes_{i=l+1}^{J} y_i \right) \left(\bigotimes_{i=1}^{l-1} y_i \right) \otimes \bigtriangledown_x y_l \right) K_{k,k^{J-1}} \\
&= J \left(x'^{\otimes (J-1)} \otimes I_k \right) K_{k,k^{J-1}} \\
&= J \left(I_k \otimes x'^{\otimes (J-1)} \right) K_{k,k^{J-1}}. \qquad (A.9)
\end{aligned}
$$

Repeating this process another $J - 1$ times we have

$$
\bigtriangledown_x^J x'^{\otimes J} = J! \prod_{j=1}^{J} \left(I_{k^{j-1}} \otimes K_{k,k^{J-j}} \right). \qquad (A.10)
$$

□

Proof to Proposition 5

$$
\begin{aligned}
&(\nabla_{\vec{Y}'} Z \otimes I_k)(I_q \otimes \nabla_X \vec{Y}) \\
&= \left[\left(\sum_{i=1}^{p} \sum_{j=1}^{q} \iota_i^p \iota_j^{q'} \otimes \nabla_{\vec{Y}'} Z_{ij} \right) \otimes I_k \right] \left[I_q \otimes \left(\sum_{u=1}^{k} \sum_{v=1}^{l} \nabla_{X_{uv}} \vec{Y}' \otimes \iota_u^k \iota_v^{l'} \right) \right] \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{u=1}^{k} \sum_{v=1}^{l} \left[\left(\iota_i^p \iota_j^{q'} \otimes \nabla_{\vec{Y}'} Z_{ij} \right) \otimes I_k \right] \left[I_q \otimes \left(\nabla_{X_{uv}} \vec{Y}' \otimes \iota_u^k \iota_v^{l'} \right) \right] \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{u=1}^{k} \sum_{v=1}^{l} \left[\left(\iota_i^p \iota_j^{q'} \otimes \nabla_{\vec{Y}'} Z_{ij} \right) \otimes I_k \right] \left[\left(I_q \otimes \nabla_{X_{uv}} \vec{Y}' \right) \otimes \iota_u^k \iota_v^{l'} \right] \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{u=1}^{k} \sum_{v=1}^{l} \left[\left(\iota_i^p \iota_j^{q'} \otimes \nabla_{\vec{Y}'} Z_{ij} \right) \left(I_q \otimes \nabla_{X_{uv}} \vec{Y}' \right) \right] \otimes \left[I_k \iota_u^k \iota_v^{l'} \right] \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{u=1}^{k} \sum_{v=1}^{l} \left[\left(\iota_i^p \iota_j^{q'} I_q \right) \otimes \left(\nabla_{\vec{Y}'} Z_{ij} \, \nabla_{X_{uv}} \, \vec{Y}' \right) \right] \otimes \left[\iota_u^k \iota_v^{l'} \right] \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{u=1}^{k} \sum_{v=1}^{l} \left[\iota_i^p \iota_j^{q'} \otimes \iota_u^k \iota_v^{l'} \right] \left(\nabla_{\vec{Y}'} Z_{ij} \, \nabla_{X_{uv}} \, \vec{Y}' \right) \\
&= \sum_{i=1}^{p} \sum_{j=1}^{q} \sum_{u=1}^{k} \sum_{v=1}^{l} \left[\iota_i^p \iota_j^{q'} \otimes \iota_u^k \iota_v^{l'} \right] \left(\frac{\partial Z_{ij}}{\partial X_{uv}} \right) \\
&= \nabla_X Z. \qquad \text{(A.11)}
\end{aligned}
$$

□

Proof to Proposition 6 (a) By induction. Put $\widetilde{x} = x - x_0$. Consider the univariate function $g(\tau) = f(x_0 + \tau\widetilde{x})$. We see by the chain rule that

$$\frac{dg(\tau)}{d\tau} = f^{(1)}(x_0 + \tau\widetilde{x})\widetilde{x} \qquad \text{(A.12)}$$

and if

$$\frac{d^j g(\tau)}{d\tau^j} = f^{(j)}(x_0 + \tau\widetilde{x})\widetilde{x}^{\otimes j} \qquad \text{(A.13)}$$

then

$$
\begin{aligned}
\frac{d^{j+1} g(\tau)}{d\tau^j} &= \nabla(f^{(j)}(x_0 + \tau\widetilde{x})\widetilde{x}^{\otimes j}) \\
&= \left((f^{(j+1)}(x_0 + \tau\widetilde{x})(I_{k^j} \otimes \widetilde{x}) \right) \widetilde{x}^{\otimes j} \\
&= f^{(j+1)}(x_0 + \tau\widetilde{x})\widetilde{x}^{\otimes j+1}. \qquad \text{(A.14)}
\end{aligned}
$$

From a Taylor expansion of $g(\tau)$ we have

$$g(\tau) = \sum_{j=0}^{J} \frac{1}{j!} g^{(j)}(\tau_0)(\tau - \tau_0)^j + \frac{1}{j!} \int_{\tau_0}^{\tau} g^{(j+1)}(t)(\tau - t)^j dt \tag{A.15}$$

where the integral form of the remainder is standard and can be confirmed by induction. Setting $\tau = 1, \tau_0 = 0$ we get the Taylor series polynomial approximation. The integral form of the remainder term is given by

$$\begin{aligned} r_J(x) &= \frac{1}{j!} \int_{\tau_0}^{\tau} g^{(j+1)}(t)(\tau - t)^j dt \\ &= \frac{1}{j!} \int_0^1 f^{(j+1)}(x_0 + t(x - x_0))(1 - t)^j dt (x - x_0)^{\otimes j+1}. \end{aligned} \tag{A.16}$$

To retrieve the Lagrange form of the remainder note that $g^{(j+1)}(t)$ is continuous and obtains its maximum ($\triangle$) and minimum ($\tilde{\triangle}$) on [0, 1]. Therefore,

$$\tilde{\triangle} \le g^{(j+1)}(t) \le \triangle, \tag{A.17}$$

$$\tilde{\triangle}(1 - t)^j \le g^{(j+1)}(t)(1 - t)^j \le \triangle(1 - t)^j, \tag{A.18}$$

$$\begin{aligned} (j+1)\tilde{\triangle} = \tilde{\triangle} \int_0^1 (1 - t)^j dt &\le \int_0^1 g^{(j+1)}(t)(1 - t)^j dt \\ &\le \triangle \int_0^1 (1 - t)^j dt = (j+1)\triangle, \end{aligned} \tag{A.19}$$

or

$$\tilde{\triangle} \le \frac{\int_0^1 g^{(j+1)}(t)(1 - t)^j dt}{j+1} \le \triangle. \tag{A.20}$$

By the intermediate value theory, there is a c such that

$$g^{(j+1)}(c) = \frac{\int_0^1 g^{(j+1)}(t)(1 - t)^j dt}{j+1} = f^{(j+1)}(x_0 + c(x - x_0))\tilde{x}^{\otimes j+1} \tag{A.21}$$

and

$$r_J(x) = \frac{1}{(J+1)!} f^{(J+1)}(x_0 + c(x - x_0))(x - x_0)^{\otimes J+1} \tag{A.22}$$

To prove (b) note that

$$\begin{aligned}\triangledown^l f_J(x) &= \sum_{j=l}^{J} \frac{1}{l!} \triangledown^l \left(f^{(j)}(x_0)(x-x_0)^{\otimes j} \right) \\ &= \sum_{j=l}^{J} \frac{1}{l!} \left(f^{(j)}(x_0) \otimes I_{k^l} \right) \triangledown^l \left((x-x_0)^{\otimes j} \right) \end{aligned} \tag{A.23}$$

and

$$\triangledown^l f_J(x_0) = \frac{1}{l!} \left(f^{(l)}(x_0) \otimes I_{k^l} \right) \triangledown^l \left((x-x_0)^{\otimes l} \right) = \frac{1}{l!} \triangledown^l \left((x-x_0)^{\prime \otimes l} \right) \triangledown^l f(x_0). \tag{A.24}$$

By Propositions 3 and 4

$$\begin{aligned}\triangledown^l f_J(x_0) &= \frac{1}{l!} l! \prod_{j=1}^{l} \left(I_{k^{j-1}} \otimes K_{k,k^{l-j}} \right) \left(I_{k^{j-1}} \otimes K_{k^{l-j},k} \right) \triangledown^l f(x_0) \\ &= \triangledown^l f(x_0). \end{aligned} \tag{A.25}$$

□

As an example of a matrix version Taylor's Theorem, consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. We refer back to the derivatives in the beginning of the "Appendix". With

$$\begin{aligned} x' &= \begin{pmatrix} x_1 & x_2 \end{pmatrix}, \\ x^{`\otimes 2} &= x' \otimes x' \\ &= \begin{pmatrix} x_1^2 & x_1x_2 & x_2x_1 & x_2^2 \end{pmatrix}, \\ x^{`\otimes 3} &= x^{`\otimes 2} \otimes x' \\ &= \begin{pmatrix} x_1^3 & x_1^2x_2 & x_1x_2x_1 & x_1x_2x_2 & x_2x_1x_1 & x_2x_1x_2 & x_2^2x_1 & x_2^3 \end{pmatrix}, \\ x^{`\otimes 4} &= x^{`\otimes 3} \otimes x' \\ &= \begin{pmatrix} x_1^4 & x_1^3x_2 & x_1^2x_2x_1 & x_1^2x_2x_2 & x_1x_2x_1x_1 & x_1x_2x_1x_2 & \cdots \\ \cdots & x_1x_2x_2x_1 & x_1x_2x_2x_2 & x_2x_1x_1x_1 & x_2x_1x_1x_2 & x_2x_1x_2x_1 & \cdots \\ \cdots & x_2x_1x_2x_2x_2^2x_1x_1 & x_2^2x_1x_2 & x_2^3x_1 & x_2^4 & & \end{pmatrix}. \end{aligned} \tag{A.26}$$

By inspection we confirm that we can write a fourth-order Taylor series approximation at a point $x = x^0$ as

$$
\begin{aligned}
f(x) &\approx f(x_0) + f^{(1)}(x_0)(x - x_0) + \frac{1}{2!} f^{(2)}(x - x_0)^{\otimes 2} + \frac{1}{3!} f^{(3)}(x - x_0)^{\otimes 3} \\
&\quad + \frac{1}{4!} f^{(4)}(x - x_0)^{\otimes 4} \\
&= \sum_{j=0}^{4} \frac{1}{j!} f^{(j)}(x_0)(x - x_0)^{\otimes j}
\end{aligned}
\tag{A.27}
$$

where

$$
\begin{aligned}
f^{(j)}(x_0)(x - x_0)^{\otimes j} = \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \cdots \sum_{i_j=1}^{k} & f^{i_1 i_2 \cdots i_j}(x_0)(x_{i_1} - x_{i_1,0}) \\
& (x_{i_2} - x_{i_2,0}) \cdots (x_{i_j} - x_{i_j,0}).
\end{aligned}
\tag{A.28}
$$

Proof of Proposition 7 (By induction.) We make repeated use of Proposition 1, in particular 1 (b) and 1 (g). For $J = 1$

$$
\begin{aligned}
\triangledown^1 A &= \triangledown (C \otimes B) \\
&= C \otimes \triangledown B + (K_{1k} \otimes I_k)(B \otimes \triangledown C) \\
&= a_0^1 (\triangledown^0 C \otimes \triangledown^{1-0} B) + a_1^1 (\triangledown^1 C \otimes \triangledown^{1-1} B)
\end{aligned}
\tag{A.29}
$$

setting $a_0^1 = I_{k^2} = I_k \otimes I_k$ and $a_1^1 = (I_k \otimes I_k)(I_{k^0} \otimes K_{k,k})$. Suppose the result holds for $J = K$ so that

$$
\triangledown^K A = \sum_{j=0}^{K} a_j^K (\triangledown^j C \otimes \triangledown^{K-j} B).
\tag{A.30}
$$

Then

$$
\begin{aligned}
\triangledown^{K+1} A &= \sum_{j=0}^{K} (a_j^K \otimes I_k) \triangledown (\triangledown^j C \otimes \triangledown^{K-j} B) \\
&= \sum_{j=0}^{K} (a_j^K \otimes I_k) \Big[(\triangledown^j C \otimes \triangledown^{K-j+1} B) \\
&\quad + (K_{k^j, k^{K-j+1}} \otimes I_k)(\triangledown^{K-j} B \otimes \triangledown^{j+1} C) \Big] \\
&= \sum_{j=0}^{K} (a_j^K \otimes I_k) \Big[(\triangledown^j C \otimes \triangledown^{K-j+1} B)
\end{aligned}
$$

$$+(K_{k^j,k^{K-j+1}} \otimes I_k)K_{k^{K-j+1},k^{j+1}}(\triangledown^{j+1}C \otimes \triangledown^{K-j}B)\Big]$$

$$= \sum_{j=0}^{K}(a_j^K \otimes I_k)\Big[(\triangledown^j C \otimes \triangledown^{K-j+1}B)$$

$$+(I_{k^j} \otimes K_{k^{K-j+1},k})(\triangledown^{j+1}C \otimes \triangledown^{K-j}B)\Big]. \tag{A.31}$$

By rearrangement of the a_j^K coefficients we have

$$\triangledown^{K+1}A = \sum_{j=0}^{K+1} a_j^{K+1}(\triangledown^j C \otimes \triangledown^{K+1-j}B) \tag{A.32}$$

where $a_0^{K+1} = a_0^K \otimes I_k$, $a_{K+1}^{K+1} = I_{k^K} \otimes K_{k,k}$ and

$$a_j^{K+1} = (a_j^K \otimes I_k) + (a_{j-1}^K \otimes I_k)(I_{k^{j-1}} \otimes K_{k^{(K+1)-j+1},k}). \tag{A.33}$$

□

Proof to Proposition 8 (By induction.) For $j = 1$, by the chain rule and application of Proposition 1 (h),

$$\begin{aligned}\triangledown_t f(t) &= (\triangledown_s g(s) \otimes I_k)(I_1 \otimes \vec{B})\\ &= (\triangledown_s g(s) \otimes I_k)\vec{B}\\ &= \mathrm{Vec}[I_k B(\triangledown_s g(s))]\\ &= B \triangledown_s g(s).\end{aligned} \tag{A.34}$$

Suppose the result holds for j: $\triangledown_t^j f(t) = B^{\otimes j} \triangledown_s^j g(s)$. Again by the chain rule and repeated application of Proposition 1, in particular 1 (h),

$$\begin{aligned}\triangledown_t(B^{\otimes j} \triangledown_t^j g(t)) &= (B^{\otimes j} \otimes I_k) \triangledown_t (\triangledown_s^j g(s))\\ &= (B^{\otimes j} \otimes I_k)[(\triangledown_s \triangledown_s^j g(s)) \otimes I_k](I_1 \otimes \vec{B})\\ &= (B^{\otimes j} \otimes I_k)\mathrm{Vec}[I_k B(\triangledown_s \triangledown_s^j g(s))]\\ &= (B^{\otimes j} \otimes I_k)\mathrm{Vec}[B(\triangledown_s \triangledown_s^j g(s))]\\ &= \mathrm{Vec}[I_k B(\triangledown_s \triangledown_s^j g(s))B^{'\otimes j}]\\ &= \mathrm{Vec}[B(\triangledown_s \triangledown_s^j g(s))B^{'\otimes j}]\\ &= (B^{\otimes j} \otimes B)\mathrm{Vec}[(\triangledown_s \triangledown_s^j g(s))]\\ &= B^{\otimes j+1} \triangledown_s^{j+1} g(s)\end{aligned} \tag{A.35}$$

noting that when $h(s)$ is $n \times 1$ and s is $k \times 1$, $\mathrm{Vec}[(\triangledown_s h(s))] = \triangledown_s h(s)$. □

Proof to Proposition 9 Since

$$\bigtriangledown (Y \otimes Z) = Y \otimes \bigtriangledown Z + (K_{s,p} \otimes I_k)(Z \otimes \bigtriangledown Y)(K_{q,t} \otimes I_l), \tag{A.36}$$

then

$$\int \bigtriangledown(Y \otimes Z)dX = \int Y \otimes \bigtriangledown Z \, dX + \int (K_{s,p} \otimes I_k)(Z \otimes \bigtriangledown Y)(K_{q,t} \otimes I_l) \, dX = 0 \tag{A.37}$$

and the result follows. □

Proof to Proposition 10 Using Proposition 9 (integration by parts)

$$\begin{aligned} \mathcal{C}_{f^{(1)}}(t) &= \int (\bigtriangledown f(x))e^{it'x} dx \\ &= -\int f(x) \bigtriangledown (e^{it'x}) \, dx \\ &= -\int f(x)e^{it'x} is \, dx \\ &= (-it)\mathcal{C}_f(t). \end{aligned} \tag{A.38}$$

Using Proposition 2 (differentiation of products),

$$\bigtriangledown((\bigtriangledown^{r-1} f(x))e^{it'x}) = e^{it'x} \bigtriangledown^r f(x) + e^{it'x}(\bigtriangledown^{r-1} f(x)) \otimes (it) \tag{A.39}$$

so

$$\int \bigtriangledown(e^{it'x} \bigtriangledown^{r-1} f(x))) \, dx = \int e^{it'x}(\bigtriangledown^{r-1} f(x)) \, dx \otimes (it) + \int e^{it'x}(\bigtriangledown^r f(x)) \, dx = 0. \tag{A.40}$$

$$0 = \int e^{it'x}(\bigtriangledown^{r-1} f(x)) \, dx \otimes (it) + \int e^{it'x} \bigtriangledown^r f(x) \, dx \tag{A.41}$$

or $\mathcal{C}_{f^{(r)}}(t) = (-it) \otimes \mathcal{C}_{f^{(r-1)}}(t)$. Repeating this $r-1$ times we have $\mathcal{C}_{f^{(r)}}(t) = (-it)^{\otimes r}\mathcal{C}_f(t)$. □

Proof to Proposition 12 Use Proposition 7 putting $A = C \otimes B = \bigtriangledown C$ and $B = \bigtriangledown \log C$ so that

$$\begin{aligned} \bigtriangledown^J C = \bigtriangledown^{J-1} A &= \sum_{j=0}^{J-1} a_j^{J-1}(\bigtriangledown^j C \otimes \bigtriangledown^{J-1-j} B) \\ &= C \bigtriangledown^{J-1} B + \sum_{j=1}^{J-1} a_j^{J-1}(\bigtriangledown^j C \otimes \bigtriangledown^{J-1-j} B) \end{aligned} \tag{A.42}$$

and

$$\bigtriangledown^J \log C = \bigtriangledown^{J-1} B = \triangle^J - \sum_{j_i=1}^{J-1} a_{j_i}^{J-1} \left(\triangle^{j_i} \otimes \bigtriangledown^{J-1-j_i} B \right). \tag{A.43}$$

Making $l = J - 1$ such substitutions we can write

$$\bigtriangledown^{J-1} B = \triangle^J + \sum_{l=1}^{J-1} (-1)^l \sum_{j_1=1}^{J_0-1} \sum_{j_2=1}^{J_1-1} \cdots \sum_{j_{l-1}=1}^{J_l-1} \left(\prod_{i=0}^{l-1} (I_{k^{J-J_i}} \otimes a_{j_{i+1}}^{J_i-1}) \right) \left(\bigotimes_{i=1}^{l} \triangle^{j_i} \right) \otimes \triangle^{J_l} \tag{A.44}$$

recalling $J_s = J - j_0 - j_1 - \cdots - j_s$, $j_0 = 0$. □

As examples let $C(t)$ be the moment generating function for a $k \times 1$ random variable having finite moments μ_J, $J \leq 4$ and consider the first four derivatives of $\log C(t)$. For $J = 2$ note that l only takes the value $l = 1$ and

$$\begin{aligned} \bigtriangledown^2 \log C &= \triangle^2 - \sum_{j_1=1}^{1} \prod_{s=0}^{0} (I_{k^{J-J_s}} \otimes a_{j_{s+1}}^{J_s-1}) \bigotimes_{s=1}^{l} \triangle^{j_s} \otimes \triangle^{J_1} \\ &= \triangle^2 - \prod_{s=0}^{0} (I_{k^0} \otimes a_1^1)(\triangle^1 \otimes \triangle^1) \end{aligned} \tag{A.45}$$

Note that $a_1^1 = K_{kk}$. Evaluating this expression at $t = 0$ we have $\bigtriangledown^2 \log C(0) = \mu_2 - \mu_1^{\otimes 2}$. For $J = 3$,

$$\begin{aligned} \bigtriangledown^3 \log C &= \triangle^J + \sum_{l=1}^{2} (-1)^l \sum_{j_1=1}^{2} \sum_{j_2=1}^{J-j_1-1} \prod_{s=0}^{l-1} (I_{k^{J-J_s}} \otimes a_{j_{s+1}}^{J_s-1}) \bigotimes_{s=1}^{l} \triangle^{j_s} \otimes \triangle^{J_l} \\ &= \triangle^3 - (I_{k^0} \otimes a_1^2)(\triangle^1 \otimes \triangle^2) - (I_{k^0} \otimes a_2^2)(\triangle^2 \otimes \triangle^1) \\ &\quad + (I_{k^0} \otimes a_1^2)(I_{k^1} \otimes a_1^1)(\triangle^1 \otimes \triangle^1 \otimes \triangle^1) \end{aligned} \tag{A.46}$$

and

$$t^{\top \otimes 3} \bigtriangledown^3 \log C(0) = t^{\top \otimes 3} \mu_3 - 3 t^{\top \otimes 3} (\mu_1 \otimes \mu_2) + 2 t^{\top \otimes 3} \mu_1^{\otimes 3}. \tag{A.47}$$

For $J = 4$,

$$\begin{aligned}
\nabla^4 \log C &= \triangle^4 + \sum_{l=1}^{3}(-1)^l \sum_{j_1=1}^{3} \sum_{j_2=1}^{J-j_1-1} \sum_{j_3=1}^{J-j_1-j_2-1} \prod_{s=0}^{l-1}(I_{k^{J-J_s}} \otimes a_{j_{s+1}}^{J_s-1}) \bigotimes_{s=1}^{l} \triangle^{j_s} \otimes \triangle^{J_l} \\
&= \triangle^4 - (I_{k^0} \otimes a_1^3)(\triangle^1 \otimes \triangle^3) - (I_{k^0} \otimes a_2^3)(\triangle^2 \otimes \triangle^2) \\
&\quad - (I_{k^0} \otimes a_3^3)(\triangle^3 \otimes \triangle^1) \\
&\quad + (I_{k^0} \otimes a_1^3)(I_{k^1} \otimes a_1^2)(\triangle^1 \otimes \triangle^1 \otimes \triangle^2) \\
&\quad + (I_{k^0} \otimes a_1^3)(I_{k^1} \otimes a_2^2)(\triangle^1 \otimes \triangle^2 \otimes \triangle^1) \\
&\quad + (I_{k^0} \otimes a_2^3)(I_{k^1} \otimes a_1^1)(\triangle^2 \otimes \triangle^1 \otimes \triangle^1) \\
&\quad - (I_{k^0} \otimes a_1^3)(I_{k^1} \otimes a_1^2)(I_{k^2} \otimes a_1^1)(\triangle^1 \otimes \triangle^1 \otimes \triangle^1 \otimes \triangle^1)
\end{aligned} \tag{A.48}$$

and

$$\begin{aligned}
t^{\top\otimes 4} \nabla^4 \log C(0) &= \mu_4 - 4t^{\top\otimes 4}(\mu_1 \otimes \mu_3) - 3t^{\top\otimes 4}\mu_2^{\otimes 2} \\
&\quad + 12t^{\top\otimes 4}(\mu_1^{\otimes 2} \otimes \mu_2) - 6t^{\top\otimes 4}\mu_1^{\otimes 4}.
\end{aligned} \tag{A.49}$$

Proof to Proposition 15 As per Proposition 6, an $(s-1)$'th order Taylor series expansion of $\mathcal{K}(t; T_N)$ at $t = 0$ yields

$$\mathcal{K}(t; T_N) = \sum_{j=0}^{s} \frac{1}{j!}\mathcal{K}^{(j)}(0; T_N)t^{\otimes j} + r_s(t), \tag{A.50}$$

$$r_s(t) = \frac{1}{s!}(\mathcal{K}^{(s)}(ct; T_N) - \mathcal{K}^{(s)}(0; T_N))t^{\otimes s}. \tag{A.51}$$

where c is between zero and one. Note that $\mathcal{K}(t; T_N) = N\mathcal{K}(t/\sqrt{N}; X)$. By Proposition 14 we have $\nabla_t^j \mathcal{K}(t/\sqrt{N}; X) = \sqrt{N}^{-j}\ \nabla_s^j\ \mathcal{K}(s; X)$. By Proposition 13 we have $\nabla^j \mathcal{K}(0; X) = i^j \kappa_j$. Substituting these results into the Taylor series and observing the first two terms are zero, we obtain the desired result. □

In anticipation of quantifying the difference between $e^{\mathcal{K}_s}$ and $\mathcal{C}_s$, temporarily define

$$c_l = \frac{1}{N^{-l/2}} \frac{\kappa'_{l+2}(it)^{\otimes(l+2)}}{(l+2)!}, \tag{A.52}$$

noting that $\mathcal{K}_s^{\dagger}(t; T_N) = \sum_{l=1}^{s} c_l$. Let $P^{\dagger}(t)$ denote a generic polynomial in t of order less than $2s$.

Lemma A1 *Put* $s_j = s - l_1 - l_2 - \cdots - l_{j-1}$ *with* $s_1 = s$. *For* $2 \leq j \leq s$

$$\mathcal{K}_s^{\dagger j} = \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_j=1}^{s_j} c_{l_1} c_{l_2} \cdots c_{l_j} + o\left(N^{-(s-2)/2}\right) P^{\dagger}(t).$$

Proof. By induction. For $j = 2$ we have

$$\mathcal{K}_s^{\dagger 2} = \sum_{l_1=1}^{s} \sum_{l_2=1}^{s} c_{l_1} c_{l_2}. \tag{A.53}$$

Note that $c_{l_1} c_{l_2} = N^{-(l_1+l_2)/2} P^{\dagger}(t)$. Therefore

$$\begin{aligned} \mathcal{K}^{\dagger 2} &= \sum_{l_1=1}^{s} \sum_{l_2=1}^{s} 1[l_1 + l_2 \leq s] c_{l_1} c_{l_2} + o\left(N^s\right) P^{\dagger}(t) \\ &= \sum_{l_1=1}^{s} \sum_{l_2=1}^{s-l_1} c_{l_1} c_{l_2} + o\left(N^s\right) P^{\dagger}(t) \\ &= \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} c_{l_1} c_{l_2} + o\left(N^s\right) P^{\dagger}(t). \end{aligned} \tag{A.54}$$

Now, suppose the result holds for some $k < j \leq s$ so that

$$\mathcal{K}^{\dagger k} = \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_k=1}^{s_k} c_{l_1} c_{l_2} \cdots c_{l_k} + o\left(N^s\right) P^{\dagger}(t). \tag{A.55}$$

Then

$$\begin{aligned} \mathcal{K}^{\dagger k+1} &= \left(\sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_k=1}^{s_k} c_{l_1} c_{l_2} \cdots c_{l_k} + o\left(N^s\right) P^{\dagger}(t) \right) \sum_{l_{k+1}=1}^{s} c_{l_{k+1}} \\ &= \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_k=1}^{s_k} \sum_{l_{k+1}=1}^{s} c_{l_1} c_{l_2} \cdots c_{l_k} c_{l_{k+1}} + o\left(N^{-s}\right) P^{\dagger}(t). \end{aligned} \tag{A.56}$$

Note that $c_{l_1} c_{l_2} \cdots c_{l_k} c_{l_{k+1}} \propto N^{-(l_1+l_2+\cdots l_{k+1})/2}$. Therefore

$$\begin{aligned} \mathcal{K}^{\dagger k+1} &= \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_k=1}^{s_k} \sum_{l_{k+1}=1}^{s} 1[l_1 + l_2 + \cdots l_{k+1} \leq s] c_{l_1} c_{l_2} \cdots c_{l_k} c_{l_{k+1}} + o\left(N^{-s}\right) \\ &= \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_k=1}^{s_k} \sum_{l_{k+1}=1}^{s_{k+1}} c_{l_1} c_{l_2} \cdots c_{l_k} c_{l_{k+1}} + o\left(N^{-s}\right). \end{aligned} \tag{A.57}$$

□

Lemma A2 $\sum_{j=0}^{s} \frac{1}{j!}\mathcal{K}^{\dagger j} = P_s(t) + o\left(N^{-(s-2)/2}\right) P^{\dagger}(t)$.

Proof. We see from Lemma A2 that

$$
\begin{aligned}
\sum_{j=0}^{s} \frac{1}{j!}\mathcal{K}^{\dagger j} &= 1 + \sum_{j=1}^{s} \frac{1}{j!}\left(\sum_{l=1}^{s} c_l\right)^j \\
&= 1 + \sum_{j=1}^{s} \frac{1}{j!} \sum_{l_1=1}^{s_1} \sum_{l_2=1}^{s_2} \cdots \sum_{l_j=1}^{s_j} c_{l_1} c_{l_2} \cdots c_{l_j} + o\left(N^{-(s-2)/2}\right) P^{\dagger}(t) \\
&\equiv P_s(t) + o\left(N^{-(s-2)/2}\right) P^{\dagger}(t). \qquad \text{(A.58)}
\end{aligned}
$$

□

Proof of Proposition 16 By the triangle inequality,

$$
\left|\int_{B_N^c(\delta)} \triangle(t) e^{-it'x} dt\right| \leq \int_{B_N^c(\delta)} \left|\mathcal{C}(t; T_N)\right| dt + \int_{B_N^c(\delta)} \left|P_s(t)\right| e^{-t'Vt/2} dt. \qquad \text{(A.59)}
$$

Since X is a continuous random variable, by Cramer's condition, for any $\delta > 0$, the first integral is $o(N^{-(s-2)/2})$. The second integral is $o(N^{-(s-2)/2})$ since $P_s(t)$ is a polynomial in t and $N^{\Delta} \int_{B_N^c(\delta)} \left|P_s(t)\right| e^{-t'Vt/2} dt = o(1)$ for any positive Δ. □

Proof of Proposition 17 To prove 17 (a) we first see, from Proposition 15, that, for some c_s in (0, 1),

$$
e^{\mathcal{K}(t;T_N) - \mathcal{K}_s(t;T_N)} - 1 = e^{r_s(t)} - 1 = r_s(t) e^{c_s r_s(t)} \qquad \text{(A.60)}
$$

where, again using Proposition 15,

$$
|r_s(t)| \leq \|R_s(t)\| \frac{(t'Vt)^{s/2}}{s! N^{(s-2)/2}}. \qquad \text{(A.61)}
$$

Note that

$$
\begin{aligned}
t'^{\otimes j} \kappa_j &= t'^{\otimes j} (V^{1/2} V^{-1/2})^{\otimes j} \kappa_j \\
&= (V^{1/2} t)'^{\otimes j} \kappa_j^* \qquad \text{(A.62)}
\end{aligned}
$$

where $\kappa_j^* = (V^{-1/2})^{\otimes j} \kappa_j$ is the j'th cumulant of the standardized random variable $Z = V^{-1/2} X$. Put $\bar{\kappa}_s^* = \max_{2 \leq j \leq s} \|\kappa_j^*\|$, which has a finite upper bound and is bounded from below by 1.

$$
\begin{aligned}
|(iV^{1/2} t)'^{\otimes j} \kappa_j^*| &\leq |V^{1/2} t'^{\otimes j} \kappa_j^*| \\
&\leq \|(V^{1/2} t)'^{\otimes j}\| \bar{\kappa}_s^* \\
&= (t'Vt)^{j/2} \bar{\kappa}_s^* \qquad \text{(A.63)}
\end{aligned}
$$

so that

$$\frac{|(iV^{1/2}t)'^{\otimes j}\kappa_j^*|}{N^{(j-2)/2}} \leq (t'Vt)\left(\frac{t'Vt}{N}\right)^{(j-1)/2}\bar{\kappa}_s^* \tag{A.64}$$

and, putting $\delta_3 = (\frac{1}{8s\bar{\kappa}_s^*})^{1/(s-1)}$ we see that, for $\|V^{1/2}t\| \leq \delta_3\sqrt{N}$,

$$\frac{|(iV^{1/2}t)'^{\otimes j}\kappa_j^*|}{N^{(j-2)/2}} \leq (t'Vt)\left(\frac{t'Vt}{N}\right)^{(j-1)/2}\bar{\kappa}_s^* \leq \frac{t'Vt}{8s} \tag{A.65}$$

and for $\|V^{1/2}t\| < \delta_3\sqrt{N}$,

$$\left|\mathcal{K}_s^\dagger(t; T_N)\right| \leq \sum_{j=3}^{s}\frac{(t'Vt)}{j!8s} \leq \frac{1}{8}t'Vt. \tag{A.66}$$

Similarly, put $\delta_2 = 1$. For all t such that $\|V^{1/2}t\| \leq \delta_2\sqrt{N}$,

$$\begin{aligned}|e^{c_s r_s(t)}| &\leq e^{\|R_s(t)\|\frac{(t'Vt)^{s/2}}{s!N^{(s-2)/2}}}\\ &\leq e^{\|R_s(t)\|t'Vt\left(\frac{t'Vt}{N}\right)^{(s-2)/2}}\\ &\leq e^{\|R_s(t)\|t'Vt}.\end{aligned} \tag{A.67}$$

Since the s'th moment of X exists, there exists a neighbourhood around zero in which $\mathcal{K}^{(s)}(t; Z)$ is continuous. Thus, for any $\epsilon > 0$, there exists a $\delta_1 > 0$ such that, for all t satisfying $|V^{1/2}t| < \delta_1$, $\|R_s(t)\| < \epsilon < \frac{1}{8}$. Choose $\delta = \min[\delta_1, \delta_2, \delta_3]$ and since $\bigtriangledown^s\mathcal{K}(t)$ is continuous we can choose any $\frac{1}{8} > \epsilon > 0$ such that for all t with $\|V^{1/2}t\| < \delta\sqrt{N}$,

$$\begin{aligned}|\triangle_1(t)| &\leq \|R_s(t)\|\frac{(t'Vt)^{s/2}}{s!N^{(s-2)/2}}e^{\{\|R_s(t)\|t'Vt|\}}e^{-\frac{1}{2}t'Vt}e^{\frac{1}{8}t'Vt}\\ &\leq \epsilon\frac{(t'Vt)^{s/2}}{N^{(s-2)/2}}e^{\frac{1}{8}t'Vt}e^{-\frac{1}{2}t'Vt}e^{\frac{1}{8}t'Vt}\\ &\leq \epsilon\frac{(t'Vt)^{s/2}}{N^{(s-2)/2}}e^{-\frac{1}{4}t'Vt}\end{aligned} \tag{A.68}$$

and

$$\begin{aligned}\int_{B_N(\delta)}|\triangle_1(t)e^{-it'x}|dt &\leq \frac{\epsilon}{N^{(s-2)/2}}\int(t'Vt)^{s/2}e^{-\frac{1}{4}t'Vt}dt\\ &= o\left(N^{-(s-2)/2}\right).\end{aligned} \tag{A.69}$$

To show 15 (b) we use a Taylor series and Lemma A2 to rewrite

$$
\begin{aligned}
e^{\mathcal{K}_s^{\dagger}(t;T_N)} - P_s(t) &= e^{c_s\mathcal{K}_s^{\dagger}(t;T_N)} \frac{\mathcal{K}_s^{\dagger s+1}(t;T_N)}{(s+1)!} + \sum_{j=0}^{s} \frac{\mathcal{K}_s^{\dagger j}(t;T_N)}{j!} - P_s(t) \\
&= e^{c_s\mathcal{K}_s^{\dagger}(t;T_N)} O\left(N^{-(s+1)/2}\right) P^{\dagger}(t) + o\left(N^{-s/2}\right) P^{\dagger}(t). \quad \text{(A.70)}
\end{aligned}
$$

where c_s is between between zero and one. As in the proof to (a), for t: $\|V^{1/2}t\| < \delta\sqrt{N}$ $|\mathcal{K}_s^{\dagger}(t;T_N)| < t'Vt/8$ so that $|(e^{c_s\mathcal{K}_s^{\dagger}(t;T_N)})e^{-t'Vt/2}| \leq e^{-t'Vt/4}$.

$$
\begin{aligned}
|\triangle_2(t)| &= O\left(N^{-(s+1)/2}\right)|P^{\dagger}(t)|e^{-t'Vt/4} + O\left(N^{-s/2}\right)|P^{\dagger}(t)|e^{-t'Vt/2} \\
&= O\left(N^{-s/2}\right)|P^{\dagger}(t)|e^{-t'Vt/4} \quad \text{(A.71)}
\end{aligned}
$$

and

$$
\left|\int_{B_N(\delta)} \triangle_2(t)e^{-it'x}\right| dt \leq \int |\triangle_2(t)|\, dt = o(N^{-s/2}). \quad \text{(A.72)}
$$

□

Proof to Theorem 18 From Propositions 16 and 17 we see that $(2\pi)^{-k}\int \triangle(t)e^{-it'x}dt = o(N^{-s/2})$. It remains to confirm the form of $f_s(x;T_N)$.

$$
(2\pi)^{-k}\int \mathcal{C}_s(t,T_N)e^{-it'x}dt = (2\pi)^{-k}\int e^{-\frac{1}{2}t'Vt}P_s(t)e^{-it'x}dt. \quad \text{(A.73)}
$$

We recognize that

$$
(-it)^{\otimes j}e^{-\frac{1}{2}t'Vt} = \mathcal{C}_{\bigtriangledown^j\phi}(t;V), \quad \text{(A.74)}
$$

that is, the jth K-derivative of the CF associated with the $N(0,V)$ density by the inversion properties of Fourier functions. Note that each of the summands in $P_s(t)$ is proportional to $(-it)^{\otimes\sum_{k=1}^{j}(l_k+2)}e^{-t'Vt/2}$. By Proposition 10 and the properties Hermite polynomials we have

$$
\frac{1}{(2\pi)^k}\int e^{-\frac{1}{2}t'Vt}(-it)^{\otimes\sum_{k=1}^{j}(l_k+2)}e^{-it'x}dt = H_{\sum_{k=1}^{j}(l_k+2)}(x;V)\phi(x;V) \quad \text{(A.75)}
$$

leading to the stated definition of $f_s(x;T_N)$. □

References

Balestra, Pietro. 1976. *La derivation matricielle, collection de l'institut de Mathematiques Economiques*. Paris: Sirey.

Barndorff-Nielson, O.E., and D.R. Cox. 1989. *Asymptotic techniques for use in statistics*. New York: Chapman and Hall.

Bhattacharya, R.N., and J.K. Ghosh. 1978. On the validity of the formal edgeworth expansion. *The Annals of Statistics* 6 (2): 434–451.

Chambers, J.M. 1967. On methods of asymptotic approximation for multivariate distributions. *Biometrika* 50 (3/4): 367–383.

Feller, William. 1971. *An introduction to probability theory and its applications*, vol. 2. New York: Wiley.

Hall, Peter. 1992. *The bootstrap and the edgeworth expansion*. New York: Springer.

Hansen, B.E. 2006. Edgeworth Expansions for the Wald and GMM Statistics for Nonlinear Restrictions, in Econometric Theory and Practice. In *Steven N*, ed. Dean Corbae. Hansen: Durlauf and Bruce E.

Holmquist, Bjorn (1996) The d-Variate Vector Hermite Polynomial of Order k, Linear Algebra and its Applications (237/238), 155–190

Jammalamadaka, S. Rao, T. Subba Rao and Gyorgy Terdik (2006) higher order cumulants of random vectors and applications to statistical inference and time series

Kollo, T. 1991. *Matrix derivatives in multivariate statistics*. Tartu: Tartu University Press.

Kollo, Tonu, and Dietrich von Rosen. 1998. A unified approach to the approximation of multivariate densities. *Scandinavian Journal of Statistics* 25: 93–109.

Kollo, Tonu, and Dietrich von Rosen 2005. Advanced Multivariate Statistics with Matrices, Springer, Berlin 93–109.

Kundhi, G., and Paul Rilstone. 2008. The third order bias of nonlinear estimators. *Communications in Statistics* 37 (16): 2617–2633.

Kundhi, G., and Paul Rilstone. 2012. Edgeworth expansions for GEL estimators. *Journal of Multivariate Analysis* 106: 118–146.

Kundhi, G., and Paul Rilstone. 2013. Edgeworth and saddlepoint expansions for nonlinear estimators. *Econometric Theory* 29: 1–22.

MacRae, E.C. 1974. Matrix derivatives with an application to an adaptive linear decision problem. *Annals of Statistics* 2: 337–346.

Magnus, Jan R. 2010. On the concept of matrix derivative. *Journal of Multivariate Analysis* 101: 2200–2206.

Magnus, Jan R., and Heinz Neudecker. 1988. *Matrix differential calculus*, 1988. New York: Wiley.

McCullagh, P. 1987. *Tensor methods in statistics*. New York: Chapman and Hall.

Phillips, P.C.B., and J.Y. Park. 1988. On the formulation of wald tests of nonlinear restrictions. *Econometrica* 56 (5): 1065–1083.

Rilstone, Paul, V.K. Srivastava, and Aman Ullah. 1996. The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics* 75 (2): 369–395.

Schild, A., and J.L. Synge. 1969. *Tensor calculus*. New York: Dover.

Taniguchi, Masanobu. 1984. Validity of Edgeworth Expansions for Statistics of Time Series. *Journal of Time Series Analysis* 5 (1):

Traat, I. 1986. Matrix calculus for multivariate distributions. *Acta et Commentationes Universitatis Tartuensis* 733: 64–84.